

No way to run a superpower

The scientific community had low expectations for the Bush administration when it was first elected. The record since 2001 shows that these expectations were justified.

From its outset, the administration of President George W. Bush drew a line in the sand on several major science-related issues, including global warming (it opposed government action to cut greenhouse-gas emissions) and ballistic-missile defence (it would be deployed, and would work). Critics, including scientists, have been shut out from providing meaningful input on these issues.

More positively, from scientists' perspective, this administration's budgets have been reasonably supportive and the president has, on the whole, appointed competent officials to key positions. Over the past two-and-a-half years, however, there has been a steady accumulation of reported attempts by the Bush administration to distort scientific evidence or rig advisory panels for political purposes. With the Republican Congress ever compliant to the administration, few questions have been raised about these complaints and the administration has been under little pressure to justify its actions.

That's why this month's report by Henry Waxman (Democrat, California), the senior Democrat on the Government Reform Committee in the House of Representatives, should be welcomed. In a relatively comprehensive and succinct 33-page document, Waxman's staff have, at the very least, put together a cogent case for the prosecution (www.politicsandscience.org; see *Nature* **424**, 715; 2003). "The Administration's political interference with science has led to misleading statements by the President, inaccurate responses to Congress, altered web sites, suppressed agency reports, erroneous international communications, and the gagging of scientists," it states.

Waxman has a good record in health and environmental issues, and his staff's detailed criticisms of administration actions deserve to be taken seriously. But the White House response has been characteristically curt, deriding Waxman as a biased observer who was "playing politics". The rest of the administration's defence has been unintentionally self-revealing. "This administration looks at the facts, and reviews the best available science based on what's right for the American people,"

White House spokesman Scott McClellan told *The New York Times*.

Administration officials have offered no point-by-point repudiation of the allegations in the report, either when they were first made or when Waxman repeated them. If they had done so, many of the allegations might have been laid to rest. There has been some political interference in agencies' decisions, and decisions have been made on the basis of political considerations rather than science. But Waxman does not establish that that this has been much more pervasive than under previous administrations.

It isn't by the prevalence of junior-level dabbling that this administration should be judged. Two things stand out in its track record on science: its handling of a few key scientific issues; and its culture of iron-clad corporate discipline, which has come into conflict with relative independence customarily enjoyed by US scientific agencies. Waxman is on his strongest ground when he attacks the impact of the latter on such nominally independent agencies as the Environmental Protection Agency (EPA) and the Centers for Disease Control and Prevention. The premature departure of Christine Todd Whitman, a moderate Republican, from the administrator's position in the EPA speaks to the curtailment of that independence under Bush.

On this occasion, Waxman doesn't dwell on the Bush administration's reluctance to engage. In 2001, the administration asked the National Academy of Sciences for an assessment of global warming that it hoped would cast aspersions on the findings of the Intergovernmental Panel on Climate Change. From the administration's point of view, the academy failed to come up with the goods, and it hasn't repeated the exercise.

On major policy issues such as global warming, ballistic-missile defence and stem-cell research, Bush committed early on to an ideologically driven approach, and has stuck to it. In an age when science pervades so many aspects of government, this is a remarkable, and remarkably ill-judged, approach to setting policy. ■

Goodbye, flat biology?

Seductive higher dimensions could consign glass dishes to history.

When the otherwise obscure Julius Richard Petri, one-time assistant to pioneering bacteriologist Robert Koch, published in 1887 his methodology for growing colonies of bacteria on a base of gelatin in flat glass dishes, he seemingly guaranteed his own immortality. The development of bacteriology and microbiology — not to mention molecular biology — would have been unthinkable without the modest Petri dish, which remains a fundamental laboratory item today.

But biologists starting to explore the merits of culturing cells in three dimensions (3-D), rather than in the flat-dish's two dimensions, have been stunned by the difference that it makes to the way the cells behave, which is much closer to their behaviour *in vivo* (see page 870). Cancer biologists find that cells can be made to switch between malignant and non-malignant states in 3-D but not in 2-D. Developmental biologists find that their cells proliferate much faster in 3-D

than in 2-D. The fat lady is not yet singing for the Petri dish, but she is certainly clearing her throat.

Culturing cells in 3-D is neither convenient nor cheap, however, and won't necessarily help to answer simple biochemical questions. But it is only a matter of time before 3-D techniques become standardized and cost-benefit ratios become irresistible in many areas of biology. Those who have experienced the merits of 3-D first-hand are convinced that any biologist studying a system where the microstructure environment is important *in vivo* — such as neurobiology, atherosclerosis or diabetes — will have to bid the Petri dish farewell in the next decade if they are to continue being taken seriously.

Awareness of the potential of 3-D tissue culture among scientists is far too low. But the benefits of the technique are so self-evident that little marketing will be needed to persuade the uninitiated to move up a dimension, just as soon as the issues of convenience are resolved. ■





Up in the air
Bioweapons treaty
remains hampered
by divisions
p865



Resistance path
Strict drug guidelines
not needed to halt
HIV mutation
p866



Heating up
Climate modellers
count the cost
of failure
p867



Publicity probe
Controversies put
scientific results
under scrutiny
p868

NASA braced for culture shock as Columbia inquiry reaches verdict

Tony Reichhardt, Washington

Six months after the loss of the space shuttle Columbia and its crew, the panel charged with investigating the accident is set to reveal its findings.

The report from the Columbia Accident Investigation Board (CAIB), due to be published on 26 August, is expected to be highly critical of the US space agency — and its implications for the future of manned space flight could be profound.

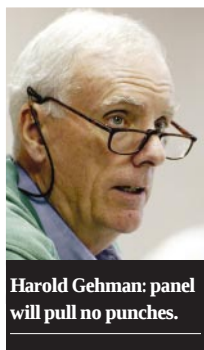
“There’s no doubt that the report will be important,” says Howard McCurdy, a space-policy expert at the American University in Washington DC. “The question is whether it will be influential, or whether everybody will graciously receive it, forget about it and go back to flight.”

Columbia disintegrated on 1 February as it attempted to re-enter Earth’s atmosphere. Uncovering the cause of the accident has been relatively straightforward, if labour-intensive. CAIB members found their ‘smoking gun’ last month following a lab test in which a chunk of insulating foam, shot from a gun at 800 kilometres per hour, blew a gaping hole in a section of the shuttle’s wing.

The test confirmed a ‘working scenario’ that the 13-member board and NASA investigators had been building for months. Columbia’s break-up was caused by just such an impact, sustained when a piece of insulating foam broke away from a fuel tank during take-off. The resulting hole allowed superheated gases to enter and fatally damage the wing during re-entry.

But other kinds of smoking gun have been harder to come by. Despite the news media’s focus on e-mails sent by low-level NASA engineers apparently warning of impending disaster before Columbia’s return, these turned out to be hypothetical scenarios rather than deeply felt concerns.

What the CAIB investigation has revealed



Harold Gehman: panel will pull no punches.



Sean O’Keefe: braced for a harsh report.



Columbia commences its ill-fated mission.

is more subtle, and perhaps more troubling — a culture at NASA that left staff staring a fatal problem in the face without noticing. Shuttle engineers had become complacent over the years about insulation foam striking the vehicle, in part because they had never analysed the risk in detail. The antiquated software used to assess the effects of the foam impact on Columbia didn’t even consider damage to material on the wing’s leading edge, where the foam actually hit. Based on this faulty assessment, flight managers accepted that there

was no danger to Columbia’s mission.

With this error in mind, the CAIB is likely to advise NASA to routinely collect far more data on the remaining shuttles’ performance and vulnerabilities than it does at present, and to make sure that potential risks come to managers’ attention. As Harold Gehman, the retired navy admiral who chairs the CAIB, said during one of the panel’s hearings: “You’ve got to start looking for trouble.”

Such a research effort could be expensive. Fixing the shuttles’ problems, while designing

R. STILES/CAIB

P. COSGROVE/AP

C. O’MEARA/AP

and building their replacement, may mean adding several billion dollars to NASA's \$15-billion annual budget, McCurdy estimates. Gehman's board won't consider the cost of its findings, but Congress and the White House will have to, especially if the CAIB recommends that NASA invest more in shuttle safety or stop flying.

Gehman is ideally positioned to issue that kind of challenge. NASA administrator Sean O'Keefe hand-picked him to head the CAIB, because of Gehman's experience in leading the investigation into the October 2000 terrorist attack on the warship *USS Cole* in Yemen, which claimed 17 lives.

Congressional critics, notably Bart Gordon (Democrat, Tennessee), a member of the House science committee, raised early doubts about the CAIB's independence, noting that it had been appointed by NASA's administrator and relied too heavily on agency staff. By contrast, Gordon pointed out, when the space shuttle *Challenger* exploded during take-off in 1986, the investigation commission was set up by the White House. In response, O'Keefe added three more panel members, all from academia, and successfully muted most of the criticism.

History repeated

The CAIB's report is set to be far more hard-hitting than the *Challenger* inquiry — partly because the institutional failings underlying the first accident seem to have been repeated. As Boston College sociologist Diane Vaughan, author of a book on the *Challenger* accident, told the panel during an April hearing: "The problems that existed at the time of *Challenger* have not been fixed."

Vaughan and other witnesses described a shuttle-engineering culture that habitually, if unintentionally, ignored trouble signs and was slow to question its own assumptions. The CAIB report will probably also take NASA to task for being too insular.

Another issue that the report will address is whether shuttle managers' bad habits are their own fault, or the result of financial and schedule pressures from above. "I think the report is going to point lots of fingers — it's not going to be a one-finger report," says McCurdy. He expects sharp criticism of Congress and of the White House Office of

Management and Budget "for deluding themselves about what's really required to run a programme of this magnitude".

Gehman's panel heard several witnesses describe long lists of safety measures that had been deferred for lack of money. In 1997, NASA estimated that it would need between \$5 billion and \$7 billion for shuttle-safety upgrades. Only a fraction of that amount was actually spent.

Testifying before the CAIB, Henry McDonald, former head of NASA's Ames Research Center in Moffett Field, California, and now an engineering professor at the University of Tennessee in Chattanooga, described the conclusion reached by a panel he chaired in 1999 to look into shuttle safety. His group found that procedures that had evolved over years to improve safety had been eroded because of a reduction in resources and staffing, he said.

Gehman's panel will consider this in identifying the root causes of the accident, and may tackle an even trickier question: how safe should we reasonably expect space flights to be? In 1995, NASA commissioned a comprehensive risk analysis for the entire shuttle system. The study, by engineering consultancy SAIC, based in San Diego, California, estimated the risk of losing a vehicle and crew at between 1 in 76 and 1 in 230. Columbia's last flight was the 88th shuttle flight after the *Challenger* disaster.

In the months since the accident, NASA managers have occasionally bristled at implications that they had been irresponsible or complacent. Phil Engelauf, a veteran shuttle-flight director and operations manager for the fateful mission, said during an emotional round-table meeting with reporters last month that the shuttle is "the most complicated machine that humans have ever built. Over time, we are going to make some human errors."

Blame culture

In his public statements, Gehman has criticized the false wisdom of hindsight. "If these flaws are out there lying around and everybody should have seen them ... tell me what the next one is, if you're so smart," he said in May. "If we as a board can't answer that question, we are very slow to sling spears at other people who also failed to answer it."

O'Keefe, meanwhile, has braced his troops for a harsh report, and agency officials have said that they will do whatever they're told. "There will be no attempt whatsoever to argue against a recommendation from the CAIB," said deputy administrator Fred Gregory, who answered reporters' questions earlier this month. Asked whether that would apply to any call to reform NASA's culture, Gregory said: "It would be difficult for me to define to you what the NASA culture is."

Whatever that culture is, attempts have



Smoking gun: a shuttle wing damaged by foam confirms the cause of Columbia's demise.

been made to reform it before — after the 1986 *Challenger* accident, most obviously, but also following the 1967 Apollo 1 fire, which killed three astronauts. An inquiry into that incident led to widespread management and engineering changes at the space agency — yet three years later, the Apollo 13 astronauts were lucky to make it back from the Moon alive, after an onboard oxygen tank exploded during the flight.

Nor have these incidents been the only motivator for reform. McDonald's safety review was initiated at NASA's own request, after an in-flight problem led the agency to halt launches and inspect wiring throughout the shuttle fleet. The review identified many of the same shortcomings that the Gehman panel will highlight next week: the waiving of safety standards; failure to update safety databases based on fresh problems; and the assumption that hardware will not fail just because it never has before.

NASA "took our report very, very seriously", recalls McDonald. The agency added 100 new inspectors, and got a commitment from President Clinton's budget office for more than \$1 billion in safety upgrades. Only months later, the incoming administration of President George Bush balked at the rising costs of the International Space Station, and told NASA to fix the problem from within its human spaceflight budget. The money for shuttle safety has since been cut drastically. One of the key figures in the budget negotiations was none other than O'Keefe, who worked at the Office of Management and Budget before taking over at NASA in November 2001.

Now, in the wake of the Columbia accident and the Gehman report, McCurdy thinks that NASA's paymasters will be forced to accept something they have long been seeking to avoid. "If you want to run the shuttle well," he says, "you've got to belly up to the fact that it costs a lot of money."

♦ www.caib.us



Piece by piece: compiling evidence on the Columbia accident has been a painstaking process.

Bioweapons initiatives bogged down in talks

Erika Check, Washington

Bioweapons experts convene in Geneva, Switzerland, this week for fresh talks on the Biological Weapons Convention (BWC), the international treaty that seeks to counter the proliferation of such weapons.

But while the United States and some of its allies will use the occasion to exhort other nations to enact stronger biosecurity laws — as called for by the treaty — their calls are likely to fall on deaf ears.

The United States and supporters such as Britain hope to use the meeting to convince other nations to enact strict pathogen-monitoring regimes and to draw up penalties for those who defy them. “We’re convinced that the best way to approach these issues of pathogen security is at the national level,” says Greg Stewart, a microbiologist seconded from the State University of West Georgia to the state department, and a US delegate to the talks.

Britain is keen that the talks are seen to produce results, meeting participants say, to demonstrate that international agreements are still relevant to global security issues. The meeting was arranged two years ago to prevent the BWC from collapsing after the United States opposed provisions to make the treaty binding on its signatory nations (see *Nature* **414**, 675; 2001). But some observers say that the best they are hoping for from this week’s meeting is that no



War of words: experts reconvene in Geneva this week to discuss controls over the spread of bioweapons.

fresh public recriminations will break out.

The talks are the first of three annual ‘experts meetings’, each to be followed by political summits, before the next major negotiations on the convention in 2006. Delegates from most of the 146 signatories to the convention, and observers from non-governmental organizations, are due to discuss penalties for nations that violate the treaty, as well as biosecurity regulations. But with the treaty itself in limbo, the United States wants to use the meeting to build pressure for biological-weapons-control measures that fall outside the scope of the convention.

Non-governmental organizations plan to use the forum to raise other issues related to arms control. Some are calling for tighter export controls and codes of conduct for governments. “We are entering a biological arms race, and European governments should be doing much more than spending two weeks in Geneva avoiding conflict,” says Jan van Aken of the Sunshine Project, a non-profit group based in Hamburg, Germany.

Others, such as Barbara Hatch Rosenberg of the Washington-based Federation of American Scientists, want the United Nations to reconstitute its weapons-inspections group as a permanent body that can be called on at a moment’s notice. “In the past the inspections took too long to get started,” she says.

A few participants hold out hope that the meeting will lend impetus to the revival of the BWC. But most think that such progress must await a change of administration in the United States. “It’s not going to happen under this administration, but these talks have been going on since the 1960s and the issue is not going to go away,” says Marie Chevrier, a political scientist at the University of Texas at Dallas who works with the Federation of American Scientists. ■

Paired-up helium atoms make it big

Geoff Brumfiel, London

Researchers at the Ecole Normale Supérieure (ENS) in Paris have coaxed two helium atoms together to create a super-sized molecule up to 100 nanometres long — ending years of argument over whether such large molecules are possible.

Helium is an inert gas whose atoms don’t like to form bonds with others. But under particular circumstances, the atoms can be brought together to form diatomic (He_2) molecules, says Jérémie Léonard, a graduate student at the ENS. Now, in work that will appear in a forthcoming issue of *Physical Review Letters*, his group has created a helium molecule far larger than any other ever made (J. Léonard *et al.* preprint at <http://arxiv.org/cond-mat/0304446>).

To create the molecules, Léonard and his colleagues cooled helium gas in a magnetic atom trap to extremely low temperatures — about 10 microkelvin — and exposed it to infrared light. The infrared photons push the electrons to one side of the positively charged helium nuclei and make the atoms

mildly attractive to each other, momentarily creating hundreds of thousands of loosely bound diatomic molecules.

Because the bond between the two component atoms is so weak, the molecules are up to a tenth of a micrometre long — thousands of times larger than most simple molecules — and only hold together for about 50 nanoseconds. But that’s still long enough for researchers to study the quantum behaviour of helium atoms and the way that they interact with others.

“There has been a lot of speculation over whether these molecules exist or not,” says Peter van der Straten, a physicist at Utrecht University in the Netherlands, who used a similar technique to make smaller helium molecules in 2000 (N. Herschbach *et al.* *Phys. Rev. Lett.* **84**, 1874–1877; 2000).

Van der Straten says he is impressed by the ENS group’s work because the team managed to detect the minute amounts of heat given off by the giant helium molecules as they broke apart. This is the first time such detection has been achieved, he says. ■



Accrued HIV evidence turns treatment dogma on its head

Erika Check, Washington

A series of studies has dispelled the widespread notion that patients who don't take every dose of their anti-HIV medication create a public-health risk by helping to nurture HIV strains that resist therapy.

The findings suggest instead that some patients who do not take all of their medicine are actually less likely to become resistant to therapy than those who adhere rigidly to their doctors' instructions.

This fresh consensus has implications for AIDS-treatment plans in poor countries, where drug companies, in particular, have often argued that HIV patients are less likely to adhere to such instructions than those in rich countries. Such 'low adherence' had been thought to encourage the success of mutations that would make the patients' viruses more resistant to treatment.

"The relationship between adherence and resistance is more complex than we previously understood," says David Bangsberg, a clinician at the University of California, San Francisco, who published a study this week showing that HIV-positive patients who were having trouble subduing the virus were more likely to develop resistance to their medications if they took most of their prescribed doses (D. R. Bangsberg *et al. AIDS* doi:10.1097/01.aids.0000076320.42412.f; 2003).

Bangsberg warns, however, that his and others' findings do not mean that patients should deliberately skip doses, because the

data still show that patients who adhere better to their drug regimens live longer. Rather, researchers say, the findings indicate that patients in developing countries should be given access to the most effective drugs early in their treatment, because this maximizes their chances of controlling the virus. Patients who control the virus successfully do not develop drug resistance.

In patients who are taking drugs but are unable to control the virus, the virus is thought to develop mutations that allow it to survive in the presence of the drugs. Such mutants do not survive in patients whose drugs keep the virus in check.

"The clear message is that in the absence of complete viral suppression, the more medication you take, the more resistance you select," says Daniel Kuritzkes, director of AIDS research at Brigham and Women's Hospital in Boston, Massachusetts, and an author of an earlier study that suggested a link between less-effective therapies and resistance (D. R. Kuritzkes *et al. J. Acquir. Immune Defic. Syndr.* 23, 26–34; 2000).

Researchers also say that data on patients in poor countries show that they are no less likely than those in rich countries to adhere to treatment. "There's no evidence to suggest that adherence is more of a challenge or resistance is more of a problem in the developing world, and there is no reason to delay the roll-out of potent anti-retroviral therapies in resource-poor settings," Kuritzkes says. ■

Water worlds make a splash as the best hope for alien life

Quirin Schiermeier, Munich

Kevin Costner's 1995 film *Waterworld* might have flopped at the box office, but researchers think that real water worlds — Earth-sized planets predominantly covered by oceans — are more likely than land-covered planets to host life.

Simple assumptions about the likely distribution of planets in the Milky Way suggest that many water worlds exist in our Galaxy, but elude existing methods of detection. "There could be as many as one billion stellar systems with potentially habitable zones," says Siegfried Franck, a geophysicist at the Potsdam Institute for Climate Impact Research in Germany.

To try to pin down the locations of planets that might host life, Franck and Manfred Cuntz, an astrophysicist at the University of Texas in Arlington, used a mathematical model to locate the 'habitable zone' of 47 UMa, a Sun-like star some 45 light years away. The pair devised equations coupling stellar age and luminosity, distance from the star, and planetary climate, to determine the chance of habitable planets existing near 47 UMa. They also calculated geodynamic constraints on the biospheres of planets that could have formed there. (S. Franck *et al. Int. J. Astrobiol.* 2, 35–39; 2003).

Earth-like planets in stable orbits in habitable zones are the most likely places to harbour life. "Earth would have a slight chance of being habitable in the 47 UMa system," says Franck, "but a water world almost entirely covered by oceans would have a better chance."

The 47 UMa system intrigues experts because the star has roughly the same mass, age and spectrum as the Sun. Moreover, it hosts two giant gas planets, analogous to Jupiter and Saturn. It is thought that such large planets help to shelter Earth from bombardment by comets and asteroids.

"Studies like this help to publicize the notion of habitable zones," says Jim Kasting, an atmospheric scientist at Pennsylvania State University. But he warns that "models of early planetary evolution are not particularly well constrained" and may not provide a reliable pointer to where inhabitable planets can be found.

NASA plans to launch two space-based telescopes, perhaps by 2013, dedicated to the pursuit of Earth-like planets, and to the analysis of their atmospheric composition. "Then the whole thing will get really exciting," says Kasting. ■



Hand of fate: HIV patients who don't stick to drug regimens seem not to foster resistant viral strains.

A. ANCETTA/GETTY IMAGES

Heatwave underlines climate-model failures

Declan Butler, Paris

Researchers are unsure whether this summer's European heatwave — which may have caused 3,000 deaths in France alone, according to government estimates — can be attributed to global warming. But there's one thing they do agree on: that it demonstrates the need for better regional climate models, to help climatologists get to grips with the processes that drive extreme weather.

Existing models of the global climate system simulate observed, long-term changes in mean temperature reasonably well, researchers say. But models of regions such as Europe are less well developed, says Simon Brown, who studies climate extremes at the UK Met Office's Hadley Centre for Climate Prediction and Research in Berkshire. Researchers blame this weakness on a lack of resources to build the models, and a lack of understanding of the processes that underlie the weather.

Global climate models are made up of data points in a three-dimensional grid from the depths of the oceans to the upper reaches of the atmosphere. The ocean, atmosphere, land surface and ice components are linked by the fluxes between them.

But even the most powerful global models are not sufficiently fine-grained to yield useful signals for regional prediction. The global models' grid points are about 2.5 degrees of latitude apart; for regional simulations, climate researchers would like to zoom down to 0.5 degrees, at least. So researchers who model local effects have to select regions of interest for higher-resolution grids, and nest these within the global models.

This is suboptimal, points out Howard Cattle, who runs the climate-change programme at the Geneva-based World Climate Research Programme. "If you force a regional model with boundary conditions from a coarse resolution model, you aren't giving it



Cold comfort: doctors in Paris try to protect elderly patients from the effects of Europe's heatwave.

information at the boundaries that is consistent with its internal dynamics," he says. "It's a 'rubbish in, rubbish out' situation."

An even bigger problem is a lack of understanding of the atmospheric processes that generate extreme climate events. The prolonged presence of a high-pressure anticyclone, which has characterized the current heatwave, is a common feature of European latitudes, says Hervé Le Treut, head of the lab of Dynamic Meteorology at the Ecole Normale Supérieure in Paris. But little is known about what controls the anticyclone's intensity, or what encourages it to hang around.

Europe's climate is particularly unpredictable, the researchers say. The El Niño effect can have a decisive effect on short-term climate trends in the United States, for example, but the influence of the Gulf Stream, say, on Europe's climate is tougher to untangle.

The fact that weather events create more

'noise' at a regional level than at a global level "makes it very hard to detect a climate-change signal" from regional data, says Ulrich Cubasch, a climate modeller at the Free University of Berlin, Germany.

The difficulty is compounded by the relatively meagre effort made to refine regional models. "Simulations by regional climate models have not been validated and analysed anywhere near as much as the global climate model simulations," says Tim Osborn, a researcher at the University of East Anglia's Climatic Research Unit in Norwich, UK.

In these circumstances, it is difficult for anyone to figure out how far this summer's heatwave in Europe reflects the wider phenomenon of global warming. But data from the heatwave will at least help researchers to test and refine their models. "We should be looking at current extremes to test the models' ability," Osborn says. ■

Europe insists fusion-project race remains wide open

Declan Butler, Paris

Officials at the European Commission have rebutted claims that an expert panel favours Spain as the potential host for ITER, the proposed international fusion experiment.

Two sites — one at Cadarache in France, the other at Vandellòs near Barcelona in Spain — are vying to be picked as Europe's candidate to host the US\$5-billion project.

The commission set up the committee of seven industrialists and scientists in May to advise Europe's ministers, who this autumn will choose just one site as the European candidate to compete with Canada and Japan to host ITER.

Reports leaked earlier this month that the panel, chaired by David King, the UK government's science adviser, preferred Vandellòs (see *Nature* 424, 606; 2003) were confirmed at the time by separate sources. But they may have reflected wishful thinking, says a source close to the panel. In reality the committee is "deeply divided", he says, and its report, scheduled for mid-summer, is unlikely to be ready until early next month.

The committee has "not been given the remit to produce a verdict", a commission official says — only to provide an impartial assessment of both sites on the basis of such

criteria as scientific and industrial expertise and site-specific costs.

The final choice of site will come not from the committee or from the commission, but from the European Council of Ministers, which is scheduled to meet at the end of September, says the committee source.

The choice of site is politically delicate, the official notes. The partners working on the research reactor — the European Union, Canada, Japan, Russia, China and the United States — hope to select a site by the end of the year. The United States has signalled its opposition to France as the site for ITER (see *Nature* 423, 211; 2003). ■

Wave of controversies spurs probe into fair presentation of results

London A series of high-profile scientific controversies has prompted Britain's Royal Society to launch a study of the way in which research results are made public.

Among the issues cited by the society is the widely derided claim that a cloned human has been born. The assertion, made last December by the Raelians, a cult that believes aliens created human life, gained widespread media coverage, but is yet to be substantiated. The society also refers to the case of disgraced physicist Jan Hendrik

Schön, which raised questions about the communication of scientific results, and particularly about peer review. A former researcher at Lucent's Bell Laboratories in Murray Hill, New Jersey, Schön was found guilty last September of fabricating and falsifying data in several prominent papers.

The Royal Society, which launched the study last week, says it will evaluate various methods for disseminating scientific results. A working group of journalists, publishers, scientists and consumer representatives has asked interested parties to submit evidence before the end of September.

NASA panel sees starry future for Hubble telescope

Washington NASA should find the money to extend the Hubble Space Telescope's life beyond 2010, says a panel of astronomers asked by the space agency to help decide the observatory's fate (see *Nature* 424, 603; 2003).

The panel, led by John Bahcall of the Institute for Advanced Study in Princeton, New Jersey, identified three options for NASA, in order of preference: have astronauts upgrade the telescope in 2005 and again in about 2010; upgrade it once in 2006, then bring it down to burn up in the atmosphere when it is no longer useful; or, if the space shuttle can no longer visit

the telescope in the wake of the Columbia accident, send an unmanned rocket to install a propulsion unit that will bring the telescope down when it ceases to function.

John Huchra, chairman of the Association of Universities for Research in Astronomy, has applauded the decision. "I am grateful that the panel established as their highest priority the maximization of science," he says. There is a catch, however: Bahcall's group says that proposals for a Hubble science programme following a second upgrade in 2010 should compete with other future missions for NASA funding. "In this way," they argue, "no approved science project would be adversely affected."

Supercomputer panel seeks winning formula

Washington The United States' policy on supercomputing technology is under scrutiny in a preliminary report from the US National Academy of Science.

Since the unveiling of Japan's Earth Simulator, which uses specially designed processors to simulate climate and weather, officials at the US Department of Energy have been calling for the United States to reduce its reliance on off-the-shelf supercomputers in favour of bespoke machines for specific problems.



Cult leader Rael's claim that his group cloned a human has angered advocates of peer review.

But in a draft report released on 12 August, an academy panel says that the government should not be too hasty in moving away from commercial machines. Instead, the report calls for a balanced investment in developing new hardware and software.

Susan Graham, a computer scientist at the University of California, Berkeley, and co-chair of the committee, says the study's results are only preliminary. "We have lots of opinions at this point," she says, "but no conclusions yet that are supported by thorough investigation." The group will present a final version of its report to the energy department in December 2004.

Sandia builds big lab for small-scale research

Washington Officials at Sandia National Laboratory in Albuquerque, New Mexico, were set to break ground this week on a half-billion-dollar facility aimed at developing advanced technologies for nuclear weapons.

The Microsystems and Engineering Sciences Applications (MESA) centre is the lab's largest-ever project. The facility, due to be completed in 2008, will house more than 600 researchers studying electronic and micro-electromechanical systems.

The US Department of Energy hopes that the centre will yield better components for

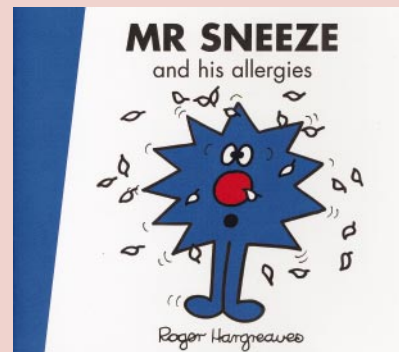
Kids' cartoon faces probe over drug tie-in deal

London Mr Sneeze, a character from the popular children's book series the Mr Men, is facing investigation from the UK Medicines and Healthcare Products Regulatory Agency over his appearance in a book promoting allergy drugs.

Pharmaceutical giant GlaxoSmithKline (GSK) commissioned the book from Adam Hargreaves, son of Roger Hargreaves, author of the original Mr Men series. *Mr Sneeze and his Allergies* follows the eponymous hero as he attempts to cope with his reaction to pollen, animals and other allergens. The final six pages contain information for parents about GSK products.

In a statement, the UK National Health Service says that the agency was unaware of the book, which may violate a 1994 law that prohibits

marketing drugs to minors. The work, the statement adds, "will be investigated urgently".



the nation's nuclear weapons. Academic scientists are also encouraged to apply to do research there.

Fermilab digs deep in bid to study neutrinos

Chicago The world's latest neutrino detector began work last week. Deep in a disused iron mine in Soudan, Minnesota, the detector, run by Fermilab, near Chicago, Illinois, will be used in the lab's Main Injector Neutrino

Oscillation Search (MINOS), which will explore different neutrino types.

For the next 18 months the detector will monitor cosmic rays that penetrate Earth, examining interactions between neutrinos and their antimatter counterparts.

Another study, set to begin in early 2005, will see MINOS scientists in Illinois firing a stream of neutrinos 700 kilometres through the ground to the Minnesota mineshaft. They hope to study how the particles' behaviour differs in space and in rock.

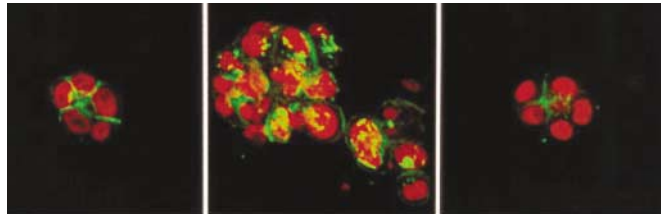
Biology's new dimension

There's a big difference between a flat layer of cells and a complex, three-dimensional tissue. But until recently, many biologists have glossed over this fact. Alison Abbott discovers what they've been missing.

Let's hear it for the humble petri dish! Many of the seminal findings in cell and molecular biology have come from cultures of cells grown cheaply and conveniently in these familiar, flat receptacles. But the limitations of considering biology in, effectively, just two dimensions are now becoming clear.

Led by cancer researchers, biologists are increasingly turning to three-dimensional cell cultures, where they are discovering patterns of gene expression and other biological activities that more closely mirror what happens in living organisms. "Scientists are starting to realize just how much a cell's context matters," says Mina Bissell, a pioneer of 3-D cell culture at the Lawrence Berkeley National Laboratory in California.

In mammalian tissues, cells connect not only to each other, but also to a support structure called the extracellular matrix



Role reversal: unlike in 2-D cultures, breast tumour cells in 3-D culture (left) that become malignant (centre) can be made to revert to their original state (right) when an antibody against β -integrin is added to the system.

(ECM). This contains proteins, such as collagen, elastin and laminin, that give tissues their mechanical properties and help to organize communication between cells embedded within the matrix. Receptors on the surface of the cells, in particular a family of proteins called the integrins, anchor their bearers to the ECM, and also determine how the cells interpret biochemical cues from their immediate surroundings.

Given this complex mechanical and biochemical interplay, it is perhaps no surprise that researchers will miss biological subtleties

if the cells they are studying grow only in flat layers. But providing an appropriate environment in which to culture cells in three dimensions is no easy matter (see 'The matrix, reinvented', below). Some researchers use simple gels consisting of collagen, whereas others make their own gels by extracting ECM material from relevant tissues. Another popular option is the commercially avail-

able Matrigel, which consists of structural proteins such as laminin and collagen, plus growth factors and enzymes, all taken from mouse tumours^{1,2}.

Culture shock

Bissell has been experimenting with 3-D culture systems for some three decades. But for years, critics argued that her methods were expensive, cumbersome and unnecessary. Their views changed after the publication of a landmark paper in 1997, in which Bissell's group showed that antibodies against a cell-

The matrix, reinvented

Interest in culturing cells in three dimensions has taken off in the past few years — but when it comes to the basic tool of the trade, most researchers are still using 1980s technology.

To grow in 3-D culture, cells need to be embedded in a structure that mimics the extracellular matrix (ECM) of structural proteins and other biological molecules found in real, living tissues. Many researchers use a material called Matrigel, a cocktail of substances extracted from the ECM of a type of mouse tumour and first described some two decades ago^{1,2}. Matrigel is a liquid below 4 °C — so cells can easily be mixed in it. Gently warming the culture then embeds the cells in the newly solidified gel.

Although Matrigel has proved effective, its inventors admit that they are surprised that it hasn't yet been superseded. "I'm absolutely shocked that there isn't really anything better," says Hynda Kleinman of the National Institute of Dental and Craniofacial Research in Bethesda, Maryland, lead author of the papers that introduced Matrigel to the world.

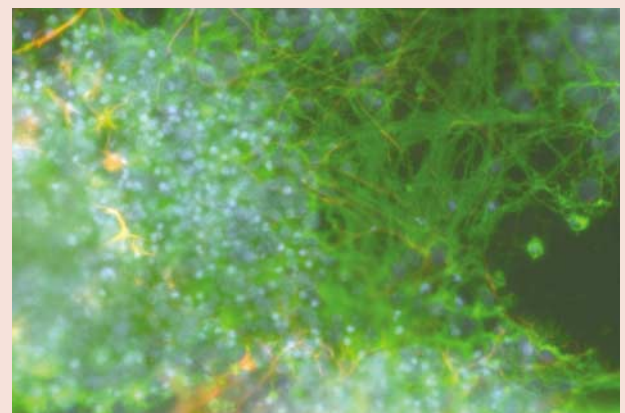
In many cases, Kleinman argues, it would be more appropriate to extract material from tissues directly relevant to the cells being studied, to provide a more suitable environment for their

growth. "People should be able to find tissue-specific matrices that would further fine-tune these investigations," she says. Indeed, some experts in 3-D tissue culture do make their own custom matrices in this way.

But in the long run, many researchers would like to get away from using materials derived from living tissues — which may vary from batch to batch and are difficult to customize for the demands of a particular experiment.

Tissue engineers who eventually want to use 3-D matrices to grow implants to graft into human patients are particularly keen on finding an alternative to materials derived from animal tissues.

Many experts predict that the future will lie in synthetic materials that can be tailor-made for specific studies. But progress has so far been slow. "It is taking us a long time to make systems that are representative and that



Growing up: neural stem cells (blue) in a 3-D protein-fibre scaffold differentiating into neurons (green) and glial cells (orange).

biologists can use easily," admits Jeffrey Hubbell, a biomedical engineer at the Swiss Federal Institute of Technology in Zurich.

Most work to date has used synthetic polymers that form 3-D matrices with micrometre-scale pores. For instance, David Mooney, a bioengineer at the University of Michigan in Ann Arbor who is interested in developing materials for biomedical tissue engineering, uses poly(lactide-



The third way: Mina Bissell says cells can behave very differently in 3-D rather than 2-D cultures.

surface receptor called $\beta 1$ -integrin completely changed the behaviour of cancerous breast cells grown in 3-D culture: they seemed to become non-cancerous, losing their abnormal shapes and patterns of growth³. This result had never been observed in 2-D cultures. Just changing the way a cell interacts with its 3-D environment, Bissell had shown, can radically alter its behaviour.

Since then, Bissell has demonstrated further important differences in the behaviour of cells grown in 2-D and 3-D cultures. For example, in the same breast-cancer system,

she has shown that antibodies against $\beta 1$ -integrin also decrease signalling by receptors for epidermal growth factor (EGF); antibodies against EGF receptors similarly depress the activity of $\beta 1$ -integrin⁴. Again, this reciprocal interaction does not happen in 2-D cultures.

Receptors for growth factors play a key role in the initial development of tumours. But this isn't the only aspect of cancer research to have benefited from the new 3-D perspective. Peter Friedl, a cell biologist at the University of Würzburg in Germany, studies metastasis — the migration of cells away from primary

tumours to cause secondary cancers around the body. Over the past couple of years, 3-D studies by Friedl's group and others have revealed unexpected subtleties in the mechanisms that cancer cells use to break out from primary tumours — clues that may help to explain the disappointing clinical performance of a promising class of cancer drugs.

Cancer cells undergoing metastasis normally cut themselves free from a tumour's ECM using protein-digesting enzymes. Yet in clinical trials, drugs that inhibit these enzymes have done little to slow the progress of cancer⁵. In his 3-D culture system, Friedl blocked the activity of the protein-chopping enzymes in two types of cancer cell, and found that the cells changed into an amoeba-like form, which could squeeze through gaps in the matrix⁶. "3-D tissue culture is really challenging our assumptions," says Friedl.

Chris Marshall, a cell biologist at the Institute of Cancer Research in London, has extended this finding, showing that the formation of amoeba-like cells depends on a particular signalling pathway in a range of different tumour cell lines. When this pathway is blocked, drugs that inhibit the protein-digesting enzymes stop the cells from moving through Matrigel⁷. This result suggests that a combination of drugs might work where inhibitors of the protein-digesting enzymes alone failed.

In another recent paper, 3-D cell culture has improved the prospect of treating cancer with gene therapy. Researchers led by Michael Korn of the University of California, San

co-glycolide), which gives a sponge-like structure with pores 100–200 μm in diameter¹¹.

But some researchers are now turning to systems based on amino acids that assemble into protein fibres of their own accord. When mixed with water, these fibres form gels with a nanoscale structure that, the researchers argue, more closely matches that of a living tissue. "Self-assembly yields nanofibres that mimic the architecture of fibrils in the ECM," enthuses materials scientist Samuel Stupp of Northwestern University in Chicago. In unpublished work, for instance, his group has used a matrix constructed from self-assembling nanofibres to coax neural stem cells into becoming neurons.

Similarly, Shuguang Zhang and his colleagues at the Massachusetts Institute of Technology's Center for Biomedical Engineering have weaved nanofibres of self-assembling peptides into a mesh with just the right porosity to slowly distribute nutrients and other necessary biological molecules to embedded liver stem cells. In this environment, the stem cells both continued to divide to reproduce themselves, and differentiated into mature liver cells¹².

Several research groups are now investigating ways to adjust the shape and surface chemistry

of the self-assembling fibres in the hope of producing gels that are better able to support cell growth. Earlier this year, for instance, Maxim Ryadnov and Derek Woolfson of the University of Sussex in Brighton, UK, tinkered with a self-assembling system to produce protein fibres that were kinked, waved or branched¹³.

But perfecting the structure of the material is only part of the battle. Doping synthetic matrices with the correct growth factors, enzymes and other molecules needed to promote the normal growth of particular cell types is far from trivial. "One of the biggest challenges is knowing what biology you need to build into your system," says Mooney.

Researchers are slowly beginning to manipulate their materials to control the biology of the cells contained within. Using different manufacturing processes to embed two different growth factors into his

matrix, for example, Mooney has succeeded in controlling the rate at which they are released by the structure¹⁴. "Most biological systems are driven by a complex combination of signals present in a defined sequence," he says.

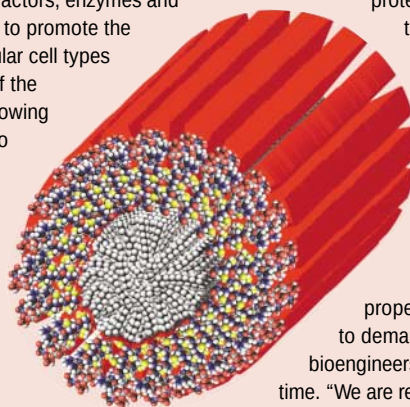
Hubbell, meanwhile, has equipped his polymer-based matrix with 'sacrificial' peptides that make it possible for cells to migrate. Cells moving through a natural ECM release

protein-digesting enzymes to cut themselves a path. Using a range of peptides in the matrix that differ in their sensitivity to degradation by these enzymes, Hubbell found that he could control the degree of movement of skin cells through his matrix¹⁵.

Given the vast range of properties that biologists are likely to demand of 3-D culture systems, bioengineers should be in for a busy time. "We are really talking about doing

hundreds, if not thousands, of different things," says

Mooney. **David Cyranoski**



Do-it-yourself: self-assembling nanofibres can be used to form an artificial matrix.

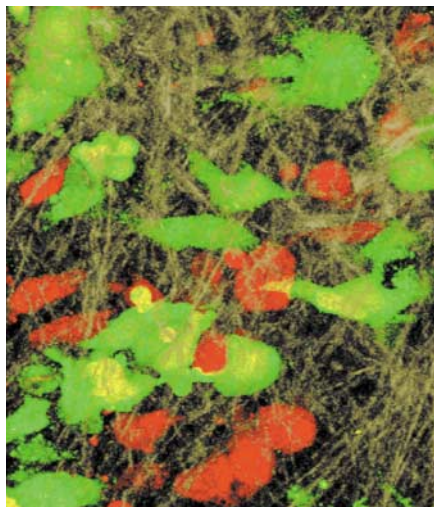
Francisco, studied the cell-surface receptors to which adenoviruses bind. In 2-D cultures, both normal and malignant breast cells had similar, high levels of the receptors. But in 3-D cultures, only malignant cells carried large numbers of the receptors⁸. Adenoviruses have been used as 'vectors' to introduce therapeutic genes into target cells, and Korn's findings suggest that they may be particularly suitable for targeting cancerous cells.

Developmental biologists are also getting in on the 3-D act. In 2001, for instance, a team led by Kenneth Yamada of the National Institute of Dental and Craniofacial Research in Bethesda, Maryland, directly compared the growth and development of fibroblasts, collagen-secreting cells that are found in many tissues, in 2-D and 3-D cultures. In three dimensions, the cells moved and divided more quickly, and assumed the characteristic asymmetric shape that fibroblasts have in living tissues⁹. "At the very least, developmental biologists who have worked with normal tissue culture will have to seriously consider comparing their results to those obtained in 3-D culture," says Yamada.

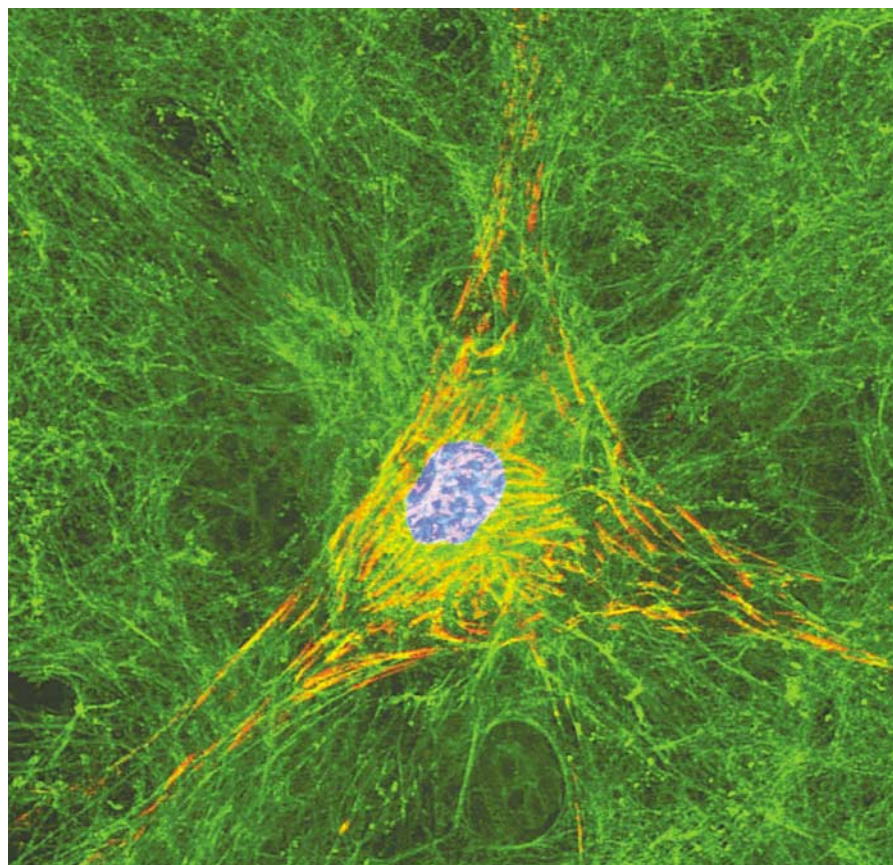
Imitating life

Some researchers are now trying to make systematic comparisons of gene activity in 2-D and 3-D cultures. In unpublished work, Linda Griffith, a bioengineer at the Massachusetts Institute of Technology, has used DNA microarrays to look at profiles of gene expression in liver cells. "Our preliminary analysis shows that the expression profile in 3-D is much closer to *in vivo* expression profiles than the profile we've seen in 2-D," she says.

If 3-D culture can provide a better model for what happens in the body, it might allow researchers to reduce their use of experimental animals — although experts stress that it is far from a complete alternative. "3-D culture



Analyses using live animals have confirmed that cancer cells (green) can escape from their location by becoming amoeba-like (red), an observation first made using a 3-D tissue culture.



Joined up: a cell in a 3-D culture forming links by means of β -integrin (orange) with the scaffolding.

will allow a lot of basic questions to be answered before having to turn to whole-animal research," says Friedl, whose work has been supported in part by a German research-ministry programme dedicated to reducing animal use. Encouragingly, when Friedl transplanted metastasizing cells into mice and used imaging techniques to track their development, they underwent the same amoeba-like morphological changes seen in 3-D culture⁶.

In October, 3-D cell culture will receive an important boost when the National Cancer Institute (NCI) in Bethesda, Maryland, launches a new section on the cellular micro-environment, which will rely heavily on 3-D studies. This programme will have an annual budget of some US\$40 million, and will include specific funding to spur the development of 3-D culturing techniques.

For Robert Weinberg of the Whitehead Institute for Biomedical Research in Cambridge, Massachusetts, the new NCI programme is a welcome development. In the 1970s and 1980s, he pioneered the study of cancer-causing genes and their associated cell-signalling pathways, mostly using 2-D cultures. "There is a whole dimension of signalling that we purposefully didn't deal with, for simplicity's sake," says Weinberg. "But now we are ready to move onto the next stage — the more complex level that 3-D culture allows."

In an article late last year, Weinberg went so far as to describe the study of cancer cells in two dimensions as "quaint, if not archaic"

ic"¹⁰. And where cancer researchers have led, he predicts, other biologists will follow.

Influential players in industry are already thinking along 3-D lines, says Mihael Polymeropoulos, chief scientific officer of Vanda Pharmaceuticals in Rockville, Maryland, and formerly head of pharmacogenetics at the Swiss-based drugs giant Novartis. "In 10 years, anyone trying to use 2-D analyses to get relevant and novel biological information will find it difficult to get funded," he predicts.

Alison Abbott is Nature's senior European correspondent; additional reporting from David Cyranoski, Nature's Asian-Pacific correspondent.

1. Kleinman H. K. et al. *Biochemistry* **21**, 6188–6193 (1982).
2. Kleinman, H. K. et al. *Biochemistry* **25**, 312–318 (1986).
3. Weaver, V. M. et al. *J. Cell Biol.* **137**, 231–245 (1997).
4. Wang, F. et al. *Proc. Natl Acad. Sci. USA* **95**, 14821–14826 (1998).
5. Overall, C. M. & López-Otín, C. *Nature Rev. Cancer* **2**, 657–672 (2002).
6. Wolf, K. et al. *J. Cell Biol.* **160**, 267–277 (2003).
7. Sahai, E. & Marshall, C. J. *Nature Cell Biol.* **5**, 711–719 (2003).
8. Anders, M. et al. *Proc. Natl Acad. Sci. USA* **100**, 1943–1948 (2003).
9. Cukierman, E., Pankov, R., Stevens, D. R. & Yamada, K. M. *Science* **294**, 1708–1712 (2001).
10. Jacks, T. & Weinberg, R. A. *Cell* **111**, 923–925 (2002).
11. Shea, L. D., Smiley, E., Bonadio, J. & Mooney, D. J. *Nature Biotechnol.* **17**, 551–554 (1999).
12. Semino, C. E., Merok, J. R., Crane, G. G., Panagiotakos, G. & Zhang, S. *Differentiation* **71**, 262–270 (2003).
13. Ryadnov, M. G. & Woolfson, D. N. *Nature Mater.* **2**, 329–332 (2003).
14. Richardson, T. P., Peters, M. C., Ennett, A. B. & Mooney, D. J. *Nature Biotechnol.* **19**, 1029–1034 (2001).
15. Lutolf, M. P. et al. *Proc. Natl Acad. Sci. USA* **100**, 5413–5418 (2003).

Take a deep breath

Want to know what's going on inside your lungs? Conventional imaging techniques are not much use, but by inhaling a magnetized gas you could get a clear picture of your airways. Erica Klarreich investigates.

James Brookeman wasn't surprised that the student volunteers taking part in his medical experiments enjoyed themselves. All they had to do was inhale xenon gas, the side effects of which are a strangely low voice and a feeling of euphoria. "They giggled quite a while after inhaling," says Brookeman, a biomedical engineer at the University of Virginia in Charlottesville. "They loved it."

While the volunteers enjoyed themselves, they were helping to change the face of medical imaging. By tweaking the magnetic properties of xenon and helium, Brookeman and others are improving magnetic resonance imaging (MRI) so that it can reveal parts of the body previously invisible to the technique.

By offering a fresh window into the lungs, these researchers should improve our understanding of diseases such as asthma and emphysema. "This is the technique of the future for looking at lung function," says Edwin van Beek, who is working on the technique at the University of Sheffield, UK. And in the long term, the experiments could lead to cheaper MRI devices and new ways of studying brain disease.

All in a row

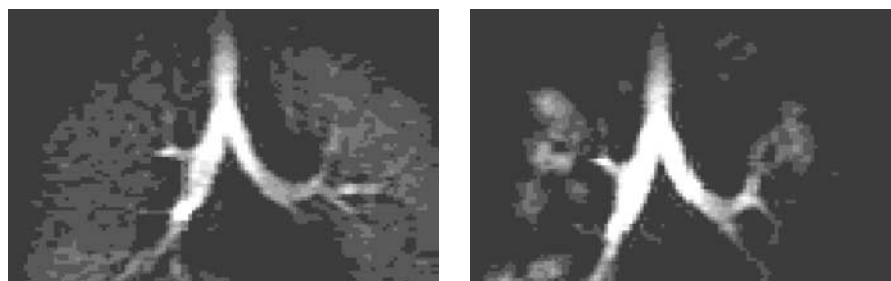
Conventional MRI makes use of the fact that hydrogen nuclei act like tiny magnets. When a patient lies inside an MRI machine, some hydrogen nuclei in water atoms in their body's tissues become polarized — they align themselves with the powerful magnetic field the machine produces. The device then delivers pulses of radio waves that temporarily tilt the nuclei, inducing an electrical current in a coil of wire in the machine. Because the nuclei in different tissues fall back into line at different rates, MRI machines can distinguish between these tissues to produce images of the brain, for example, or the kidneys.

But not all organs can be imaged in this way. The lungs contain little water, and so have been off-limits to MRI scans. Other imaging techniques, such as positron-emission tomography (PET), which tracks the movement of radionuclides around the body, can image the lungs. But PET scans expose the patient to radiation, making them impractical tools for studying diseases, such as asthma, which would require patients to have regular scans.

Such problems led Brookeman and others



Magnetic resonance imaging offers an insight into the body, but until now it could not show the lungs.



Clear view: normal lungs (left) and a simulated asthma attack taken using gas-based imaging.

to explore an alternative technique: filling the airways of the lungs with a gas that can respond to MRI. The idea dates back to the early 1990s, and is the brainchild of William Happer and Gordon Cates, then both at Princeton University in New Jersey. The two physicists were studying the process of hyperpolarization, which involves using a polarized beam of laser light to align the nuclei of

gaseous atoms. Their work wasn't aimed at medicine — hyperpolarized helium-3 gas is used as a target in particle-scattering experiments — but they realized that a hyperpolarized gas should react to radio pulses in a similar way to aligned hydrogen nuclei.

In 1994, Cates and others described the first image taken using hyperpolarized xenon-129 gas in a lung, in this case from a

dead mouse (M. S. Albert *et al. Nature* **370**, 199–201; 1994). The resulting lung image was a spectacular success, and was soon followed by *in vivo* images of human and mouse lungs (see, for example, M. Ebert *et al. Lancet* **347**, 1297–1299; 1996).

In theory, the technique offers significant benefits. A much greater proportion of nuclei are aligned in hyperpolarized gases than in water exposed to the magnetic fields in MRI, so the signals produced using the gas are about 100 times stronger. As a result, a hyperpolarized gas can image an entire lung in the few seconds it takes a patient to inhale, hold their breath, and exhale — a marked improvement over the hour or so it typically takes to get an image using conventional MRI. By taking images as a patient breathes in and out, the new technique can also produce dynamic images of airflow in and out of the lungs.

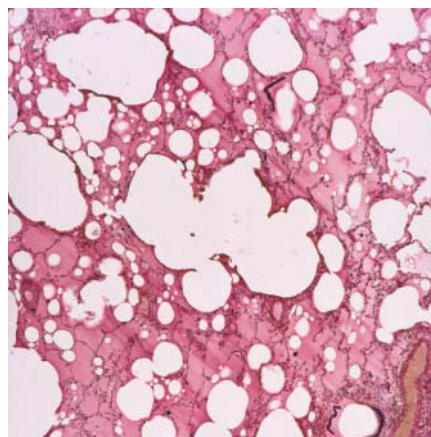
Polar vision

The treatment and diagnosis of asthma could be one area to benefit. In an asthma attack, branches of the tree-like structure of the lungs' airways either narrow or shut down. This June, Brookeman and his team described how they used hyperpolarized helium to detect which portions of the lungs were cut off during an asthma attack — areas where gas flow was reduced or eliminated showed up darker in the images (S. Samee *et al. J. Allergy Clin. Immunol.* **111**, 1205–1211; 2003). "Parts of the lung totally closed; other parts remained normal," says Brookeman. "Before, it wasn't clear whether the narrowing happened everywhere or not."

Hyperpolarized gas MRI should also be valuable for testing the efficacy of asthma therapies, says Mitchell Albert, a radiologist who works on hyperpolarization at Brigham and Women's Hospital in Boston. "We'll be able to figure out if one treatment is better than another, what is the optimal time for medication, how frequently patients should use their medication, and so on," he says.

Another disease that may reap the benefits offered by hyperpolarized gas MRI is emphysema. This condition affects the walls of the small air sacs, called alveoli, that act as the interface between the lungs and the bloodstream. The walls break down, merging alveoli together to form much larger air sacs. This greatly reduces the surface area available for oxygen transport into the blood, leaving the sufferer fighting for breath.

Dmitriy Yablonskiy, a physicist at Washington University in St Louis, Missouri, is part of a team investigating the rate at which hyperpolarized gas diffuses through the lungs of emphysema patients. In a healthy lung, diffusion in the alveoli is relatively slow, as gas atoms can't move far before bumping into the edge of a sac. Diffusion in emphysema patients is faster, because the large alveoli



Gas-based imaging may allow rapid diagnosis of emphysema (below), in which the lungs' small air sacs become enlarged (above).



give the gas more room to move. "The diffusion coefficient is four times larger in emphysema patients than healthy people," says Yablonskiy.

Such discoveries could help to improve diagnosis. Brookeman's team is using diffusion measurements to study smoking-induced emphysema. He notes that patients do not usually start showing symptoms of the disease until after 20 pack-years of smoking (a pack-year is equivalent to smoking a packet of 20 cigarettes a day for a year). "But with a diffusion map of the lungs, it looks as though we can spot the effects of smoking after five pack-years," he says. "We can tell people that they should quit because their lungs are starting to deteriorate, even though they have no symptoms."

Shadowy figures

Armed with hyperpolarized gas and oxygen, researchers are optimistic that they will also be able to image blood clots in the lungs. Oxygen increases the speed at which hyperpolarized helium and xenon lose their polarization. Blood clots prevent oxygen from being absorbed in the lungs, so if a patient inhales oxygen, it will linger in the vicinity of the clot. If the patient then inhales hyperpolarized gas, areas where the oxygen hasn't been absorbed will give a much weaker MRI signal than other areas, showing up as a dark area on the scan.

But the lungs may be just the beginning — xenon-129 has an added advantage that

should allow it to image other organs. Unlike helium, xenon dissolves in the blood, reaching the brain and other organs within seconds of being inhaled. Albert is involved in a project to investigate whether xenon could be used to study brain plaques, lumps of tissue that build up in the nervous system and are associated with diseases such as multiple sclerosis. Albert hopes that xenon imaging will provide a way to distinguish between plaques caused by multiple sclerosis and those caused by other factors such as blood-flow problems.

Taking images with hyperpolarized gas MRI could be relatively cheap, says John Owers-Bradley, a physicist at the University of Nottingham, UK. Conventional MRI scanners use superconducting magnets to align hydrogen nuclei, and these need to be cooled by liquid helium. But in the new technique, the gas is already polarized, so a small magnetic field is needed to generate the final image. "We've produced good pictures with fields just one-tenth the usual strength," says Owers-Bradley. And compared with hydrogen atoms, which move out of alignment within seconds of a magnetic field being removed, isolated samples of hyperpolarized helium and xenon can remain polarized for more than a day. This means that the gas for a lung MRI can be prepared in advance, even in a separate facility, making it easier for small hospitals to adopt the technology.

The X factor

So far, most human trials of the technique have used helium-3 rather than xenon-129, mainly because helium is easier to hyperpolarize than xenon. But techniques for improving xenon polarization are currently being developed. This is important, as helium-3 does not occur in large quantities naturally. In fact, the only abundant source of the gas is nuclear-weapons programmes — helium-3 is a decay product of tritium, which is used in the trigger of hydrogen bombs. "In the early days, we'd usually have to deal with some Russian general who wanted the money put in a Swiss bank account," says Brookeman.

The US and Russian governments treat the gas as a strategic material, so although several companies are now buying helium-3 from governments and selling it to MRI researchers, no one is sure just how large the stocks are worldwide. The supply is probably enough for about 10 years' worth of imaging, Happer guesses. Xenon-129, by contrast, is relatively abundant: it makes up nearly one-quarter of the xenon in the atmosphere.

With cheap and high-resolution images on the horizon, hyperpolarization looks set to boost medical imaging, transforming the way in which we study certain diseases. "I'm convinced it's going to be of great benefit," says Brookeman. His giggling students, it would seem, have spent their time wisely. ■

Erica Klarreich is a freelance writer in Berkeley, California.

Need for a court to rule on patents and misconduct

Scientific affairs are too complex to be solved locally, says a researcher cleared of fraud.

Sir — Disputes over promotion and control of resource allocation occur in every lab. Fortunately, they rarely end up with the destruction of a solid research team as reported in your News story “Tribunal clears obesity researcher of fraud” (*Nature* **424**, 6; 2003). If all participants had only been willing to follow a few simple rules, the situation could have been avoided.

First, the details of the allegation should have been brought forward immediately. Because these were kept secret for a long time, a simple, transparent scientific debate became impossible. Instead, two adversarial sides emerged. It took five years for a prosecutor and a judge to understand and rule on the matter. French justice was the sole institution able to get to the bottom of this longstanding problem because the law provided it with the full power of disclosure.

Whistleblowers should not be allowed anonymity. The remote possibility of retaliation does not justify secrecy, and protection by the authorities can easily be guaranteed. The openness of the procedure

will itself shield the accuser from the accused. Bringing forward an accusation that jeopardizes someone's career requires courage: both the accused and the whistleblower should take responsibility for their actions.

The main lesson to be learned from the whole affair is that any attempt by anyone to maintain any part of an investigation of this type in secrecy is the truth's biggest enemy. Even if the inevitable invasion of privacy is sometimes hard to bear, it is the only way.

Second, although I strongly advocate complete openness when confronted by the inquiry procedure, I also believe that disclosure to the public by journalists does more harm than good. What would have happened had you decided not to publish your stories about this affair? Rennes University would have pursued the case, the judicial inquiry would have been completed, my name would have been cleared and I believe Inserm Unit 391 would have been able to pursue its work. This is not to say that journalists should not report on cases of misconduct. But they should

wait until they have definitive evidence. Scientific articles based on circumstantial evidence and not peer-reviewed are not published, because a general audience cannot critically evaluate the data. The same rule should apply to journalism.

Finally, matters such as this one cannot be handled at the local level. I believe the best way forward would be for a court of scientific affairs to be established: possibly at European level. Although there are only a few cases of scientific misconduct, there are many patent disputes that require detailed understanding of science. With the growing industrialization of biology more such disputes are likely to happen. Science is now mixed with large economic and political interests requiring a powerful court of arbitration. Such a court should be staffed by people with impeccable scientific and legal credentials, but with no direct involvement in scientific research policy.

Bernard E. Bihain

*Médecine et Thérapeutique Moléculaire (MTM),
15 rue du Bois de la Champelle,
54500 Vandœuvre-les-Nancy, France*

Haeckel's literary hopes dashed by materialism?

Sir — The renewed debate (see, for example, ref. 1) about whether the German zoologist Ernst Haeckel ‘doctored’ his embryo drawings to fit his biogenetic law — ontogeny recapitulates phylogeny — has led to intensified research into other aspects of Haeckel's life and work. One example is the strange story of when he thought he had been awarded the 1908 Nobel Prize in Literature.

Haeckel was very disappointed when he heard that this prize had gone to the German philosopher Rudolf Eucken. French and Italian newspapers had announced that Haeckel was to be given the prize, and he had received telegrams and postcards congratulating him. Haeckel thought he deserved the prize, and wrote to a friend, the publisher Wilhelm Breitenbach, on 30 November 1908: “If I were to get the Nobel prize (which in view of my 50 years of work and according to the often expressed views of colleagues might be justified!) I would donate the money to the Phyletic Museum.”

It was particularly irritating for Haeckel to be beaten by Eucken. On 29 December that year, Haeckel wrote (see page 214 of ref. 2) to his friend and biographer, the popular science writer Wilhelm Bölsche: “I heard from Stockholm that there had

actually been a kind of competition in the ‘Nobel commission’ between myself and my colleague Rudolf Eucken. But the latter won as an advocate of idealism and a priest of the ‘higher spiritual world’, while I as advocate of materialism and slave to the ‘lower Nature’ had to lose. Eucken is a popular rhetorician and promoter of the Christian religion, but until now he has not brought any new ideas into philosophy.”

Haeckel's belief that he was nominated for a Nobel Prize in Literature is incorrect (see, for example, ref. 3 — research by U.H. and L.O. in Swedish archives was supported by a grant from the Royal Swedish Academy of Sciences' Centre for the History of Science). Among the 16 nominees, the top candidates initially were the Swedish novelist Selma Lagerlöf and the English poet Algernon Charles Swinburne, but the Nobel committee could not make up its mind between these two. Eucken was suggested as a compromise solution by Vitalis Norström, professor of philosophy at Gothenburg University, who admired Eucken's philosophical writings⁴. These were seen as consistent with the terms of Alfred Nobel's will directing that the literature prize should go to a work written “with an idealistic tendency”.

The award of the literature prize to Eucken has been called “the biggest *faux pas*”

in the history of the Nobel prize (see page 63 of ref. 5). Haeckel's conviction that his materialism was unpopular among leading members of the Swedish Academy receives support in a letter dated 27 November 1908 from the historian Harald Hjärne, director of the academy, to the poet and academy member Esaias Tegnér. Hjärne wrote that Eucken was needed “as a counterweight to the demonstrations in support of his Jena colleague Haeckel” (see page 181 of ref. 6) during Uppsala's 1907 bicentennial celebrations for Linnaeus, when a lecture by Haeckel had been enthusiastically received.

**Uwe Hossfeld*, Rosemarie Nöthlich*,
Lennart Olsson†**

**Institut für Geschichte der Medizin,
Naturwissenschaft und Technik,
Ernst-Haeckel-Haus, Friedrich-Schiller-Universität,
Berggasse 7, D-07745 Jena, Germany*

*†Institut für Spezielle Zoologie und Evolutionsbiologie
mit Phyletischem Museum, Friedrich-Schiller-Universität, Erbertstr. 1, D-07743 Jena, Germany*

- Kemp, M. *Nature* **395**, 447 (1998).
- Nöthlich, R. *Ernst Haeckel—Wilhelm Bölsche: Briefwechsel 1887–1919* (VWB, Berlin, 2002).
- Svensén, B. *Nobelpriset i Litteratur. Nomineringar och utlåtanden 1901–1950. Del. I: 1901–1920* (Norstedts, Stockholm, 2001).
- Eucken, R. *Grundlinien einer neuen Lebensanschauung* (Veit, Leipzig, 1907).
- Lång, H. *De litterära Nobelprisen 1901–1983* (Bra Böcker, Höganäs, 1984).
- Espmark, K. *Det litterära Nobelpriset. Principer och värderingar bakom besluten* (Norstedts, Stockholm, 1986).

Who said 'helix'?

Right and wrong in the story of how the structure of DNA was discovered.

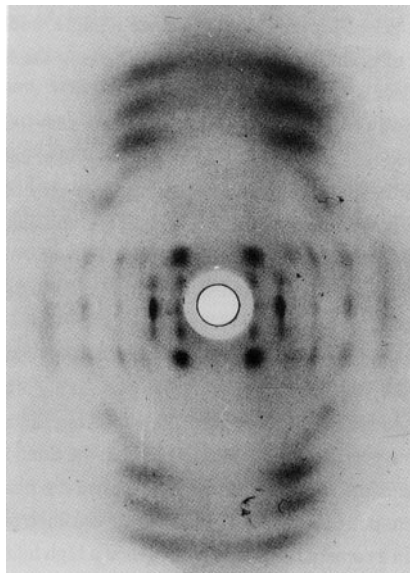
Watson Fuller

The celebrated model of DNA, put forward in this journal in 1953 by James Watson and Francis Crick, is compellingly simple, both in its form and its functional implications (see www.nature.com/nature/dna50). At a stroke it resolved the puzzle inherent in the X-ray diffraction photograph (see right) shown by Maurice Wilkins at a scientific meeting in Naples in the spring of 1951. This was the pattern that so excited Jim Watson, who, in *The Double Helix*¹, wrote: "Maurice's X-ray diffraction pattern of DNA was to the point. It was flicked on the screen near the end of his talk. Maurice's dry English form did not permit enthusiasm as he stated that the picture showed much more detail than previous pictures and could, in fact, be considered as arising from a crystalline substance. And when the structure of DNA was known, we might be in a better position to understand how genes work."

DNA structure investigations

This pattern had been recorded at King's College London by Raymond Gosling from a bundle of thin fibres drawn by Wilkins from DNA provided by Rudolf Signer in early 1950. Alec Stokes, also at King's, pointed out that apart from strong 3–4-Å features at the top and bottom of the pattern, there was no diffraction close to the meridian, indicating that DNA had a helical structure². From measurements by Gosling, the vertical separation of the rows of spots (layer lines) could be identified with a helix pitch of 28 Å, and the lateral spacing of spots with helices packed together like 20-Å diameter cylinders. The strong 3–4-Å features had previously been observed in Bill Astbury's laboratory in Leeds and had been attributed to diffraction from planar bases in DNA, stacked like "a pile of plates"³.

Despite the relatively primitive X-ray equipment of the time, by the end of 1950 the group at King's had not only demonstrated that DNA has a highly regular structure, but had extracted from the diffraction data important clues about the shape of the molecule and the dimensions of features in it. Wilkins and his colleagues had good cause for optimism, and it was expected that the arrival of Rosalind Franklin from a Paris laboratory, where she had already successfully exploited X-ray diffraction techniques, would further strengthen the group. Much has been written about the clashes, particularly with Wilkins, that marked Franklin's two years at King's (see, for example, refs 4, 5). Here I aim to provide a more accurate



A crystalline X-ray diffraction pattern of DNA, taken in June 1950 by Raymond Gosling, from a multifibre specimen made by Maurice Wilkins from DNA supplied by Rudolf Signer.

A Raymax X-ray tube was used with a Unicam camera, filled with hydrogen (90% relative humidity). In Wilkins' words: "now it was really obvious — genes had a crystalline structure".

perspective on the scientific role of Wilkins than is often apparent in popular narratives of the discovery — and even among some sections of the scientific community.

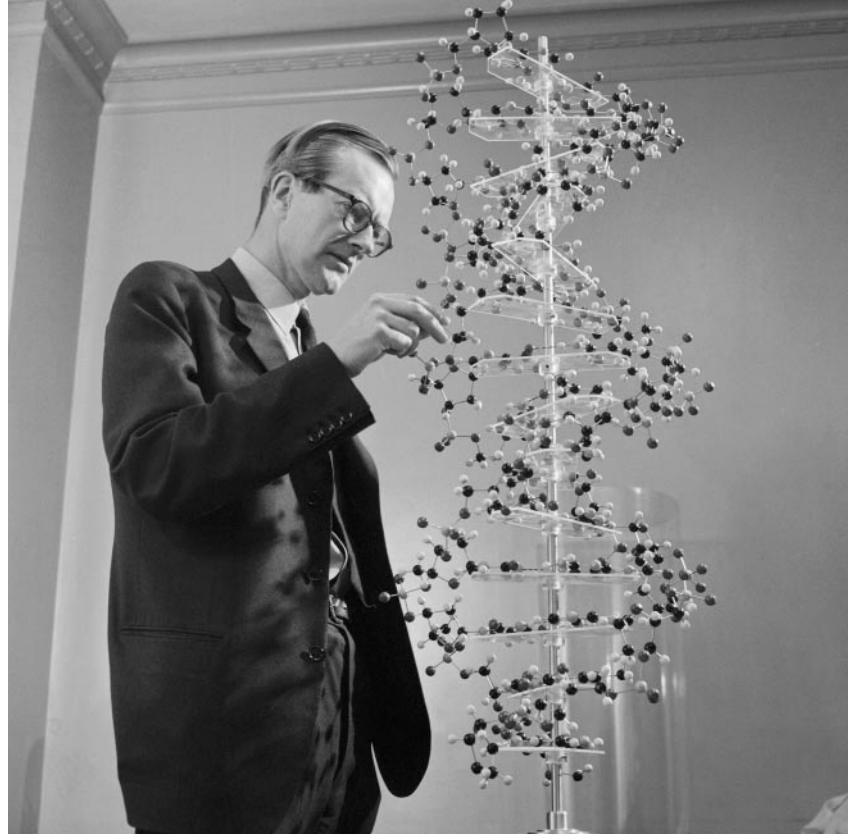
Certainly Franklin, an established researcher, had good reason to feel that she had been misled by a letter from the head of the King's laboratory, John Randall, about the degree of independence she could expect in her X-ray diffraction studies of DNA. Equally, Wilkins could feel outraged at being abruptly excluded — with no prior notice — from the very aspect of the DNA work on which he and Gosling had made such excellent progress. Randall, with his experience of directing an industrial research laboratory and of sharply focused, technological war-directed research, might have felt that the *laissez-faire* ethos of university research ought to be superseded. Moving staff around like pieces on a chess board could expedite the results that would justify generous levels of public funding. But almost a year before the end of her three-year fellowship, Franklin, to widespread relief at King's, left to join J. D. Bernal's research group in Birkbeck College. As a result, Randall must have seriously questioned his own actions and have wondered whether, but for the conflicts engendered by the letter he had written to her

in Paris, the double-helical structure of DNA might not have been discovered in London rather than in Cambridge. In fairness to Randall, it was his energy, enterprise and vision in establishing the King's laboratory that allowed the experimental work that stimulated the discovery to take place.

The proposed double-helical model for DNA is commonly described as the most significant discovery of the second half of the twentieth century. Inevitably, the contributions of the principal protagonists have been subjected to minute scrutiny. Crick, Franklin, Watson and Wilkins have all endured hostile criticism and snide disparagement of their roles in the story. Franklin has loyal, influential and persistent champions, and in particular has had her reputation boosted, mainly at Wilkins' expense. Some have gone so far as to ask whether, if Franklin, who died at the age of 37 in 1958, had lived, she would have shared the Nobel prize for the discovery of the structure of DNA. As the prize is never shared by more than three people, the implication is that Wilkins would have been excluded.

Humidity-driven transitions

By far the most significant contribution by Franklin while at King's was the exploitation of the fine-focus X-ray generator developed by Werner Ehrenberg and Walter Spear at Birkbeck College, London, previously obtained by Wilkins, and a commercial X-ray microcamera in experiments in which the humidity of the specimen environment was accurately controlled. Although in their earlier experiments Wilkins and Gosling had realized serendipitously that DNA should be studied in a moist environment (famously maintained by stretching a condom around the camera collimator), they had not achieved precise control nor been able to sustain very high humidities reproducibly. Franklin's improvements allowed her and Gosling to undertake a systematic study of the manner in which the X-ray diffraction from DNA changed with the water content of the fibre and to obtain patterns with much sharper diffraction maxima⁶. They identified two forms, 'A' (a low-humidity form) and 'B' (high humidity), together with intermediate patterns of A mixed with B. Both forms had been observed by previous workers, notably the image shown above (the A form) recorded by Wilkins and co-workers, as well as less-good B-form patterns by them and in Astbury's laboratory⁷. The analysis of the improved A and B patterns by Franklin and Gosling had important consequences, positive and negative,



Maurice Wilkins in 1962 posing for reporters following the announcement that he, Watson and Crick had won the Nobel prize for their work on the elucidation of the structure of DNA.

both for the work at King's and for the development of the double-helical model.

A typical crystal or fibre studied by X-ray diffraction contains billions of molecules; the general assumption is that the purer the sample and the greater the degree of regularity with which the molecules are arranged, the sharper will be the spots in its diffraction pattern. The position of the spots reflects the disposition of the molecules in the crystal or fibre, and the overall intensity variation across the pattern, which can be imagined as being 'sampled' at the points at which the spots appear, is determined by the overall shape of the molecule. Determination of the arrangement of the molecules from the position of the diffraction spots was relatively straightforward, but determination of the shape of the individual molecule from the relative intensities of these spots required a combination of insight, experience and good fortune.

Franklin and Gosling concentrated their efforts on analysis of the A pattern which, because it had sharp spots across its whole extent, was designated as crystalline, rather than on the B pattern, which was designated semi-crystalline because all of the sharp spots were confined to the centre with continuous streaks in the outer regions. Although they pursued their goal with high analytical skill, methodological innovation and a willingness to undertake arduous calculations, the endeavour was unsuccessful. Franklin had been adamant that the investigation should proceed by a detailed analysis of the X-ray data rather than a more intuitive

IT IS WITH GREAT REGRET THAT WE HAVE TO ANNOUNCE THE DEATH, ON FRIDAY 18TH JULY 1952 OF D.N.A. HELIX (CRYSTALLINE) DEATH FOLLOWED A PROTRACTED ILLNESS WHICH AN INTENSIVE COURSE OF BESSELIER'S INJECTIONS HAS FAILED TO RELIEVE. A MEMORIAL SERVICE WILL BE HELD NEXT MONDAY OR TUESDAY. IT IS HOPED THAT DR. M.H.F. WILKINS WILL SPEAK IN MEMORY OF THE LATE HELIX R.E. Franklin Rosalind

In Memoriam card sent by Rosalind Franklin and Raymond Gosling to Maurice Wilkins in July 1952 (published in ref. 4). The "INTENSIVE COURSE OF BESSELIER'S INJECTIONS" refers to the expression derived by Alec Stokes for diffraction from helical molecules in terms of cylindrical Bessel functions.

approach, building on Stokes' inference that the overall distribution of diffracted intensity in both the A and B patterns, in particular the lack of intensity along the meridian, strongly suggested that the structure of DNA in both forms was helical. Opinions at King's became increasingly polarized, with Franklin more and more strongly identified with an anti-helical view, expressed most triumphantly in the *In Memoriam* card distributed by Franklin and Gosling in July 1952 announcing that their analysis of the A form heralded the death of the helix (see above). However, Aaron Klug, from his access to Franklin's notebooks and draft papers, has shown that Franklin was not as anti-helical as was widely believed in the months leading up to the discovery of the double helix, and that indeed

she was working on the assumption that the B form at least was helical^{8,9}.

In the much improved B-type pattern obtained by Franklin and Gosling in May 1952, the helical features are very clear. It was this pattern that Gosling passed to Wilkins, perhaps naively, but totally in character given his strong commitment to sharing scientific information, showed it to Watson on his crucial visit to King's at the end of January 1953. Watson immediately recognized this as the clearest evidence yet for a helical structure. Accounts vary as to the details of the B pattern Watson took back to Cambridge. Some have read signs of unease into the Watson and Crick paper of 1953, in that it seems evasive about the extent to which they used information from the King's laboratory to define their model, in particular, the helix pitch of 34 Å with 10 residues per turn and the strongly scattering phosphate groups at a radius of 10 Å.

Watson and Crick, as Watson notoriously stated in *The Double Helix*, also had the benefit of seeing a report that was submitted by King's scientists to the Medical Research Council (MRC). This was given to Max Perutz in the Cambridge laboratory in his role as a member of a review committee. Views differ over the degree of confidentiality of this document and whether Perutz behaved properly in showing it to his colleagues, including Crick. In Perutz's defence, the point has been made that one of the purposes of these reports was to make the various MRC laboratories aware of what the others were doing. Recollections differ on how much of the information in this report (written in early September 1952 for a committee visit in December 1952, and shown to Crick in February 1953) was included in a seminar by Franklin at King's in November 1951 (to which Crick was invited but sent Watson). This report included the determination by Franklin and Gosling of the symmetry of the unit cell of the A form. This was to prove, in the hands of Crick, crucial information, as it implied that DNA had two-fold symmetry about an axis perpendicular to the helix axis, and therefore, if it was a two-stranded structure, that the direction of one chain should be opposite to that of the other.

Whether because of discomfort over the less than customary channels for the flow of information from King's to Cambridge or not, the double-helical model is presented in the epoch-making paper by Watson and Crick more as a revelation to prepared minds than the product of a series of logical steps from crucial experimental data. Thus Watson and Crick were at pains to stress in their paper the importance of further X-ray work to confirm their hypothesis. Over the next decade, Wilkins and his co-workers indeed obtained X-ray diffraction patterns that were much better defined, and developed

new techniques for their analysis^{10–12}.

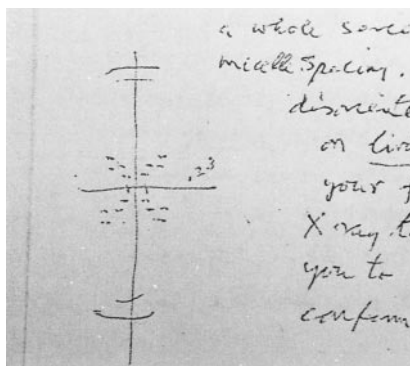
Two papers, one by Wilkins, Stokes and Herbert Wilson, the other by Franklin and Gosling, were published alongside that of Watson and Crick (see www.nature.com/nature/dna50), showing that the Watson and Crick model was consistent with their X-ray diffraction patterns. The paper by Wilkins *et al.* also referred to A- and B-type diffraction patterns from various sources and notes the “most marked” correspondence between their presentation of the diffraction expected from the double-helical model and “the exceptional photograph obtained by our colleagues, R. E. Franklin and R. G. Gosling”. For their part, Franklin and Gosling noted that Stokes and Wilkins were the first to propose helical structures as a result of direct studies of nucleic-acid fibres. Indeed, in a letter to Crick in early 1952, Wilkins had already sketched out the diffraction pattern from oriented sperm heads showing B-type diffraction, with the cross-like intensity distribution characteristic of a helix on layer lines in the centre of the pattern and meridional diffraction at the top and bottom (see above right).

Molecular model building

Crucial for the success of Watson and Crick's approach had been the use of molecular models based on knowledge of chemical bonding determined from stereochemistry and X-ray single-crystal studies of molecular components. The models allowed rotation about single bonds in the chemical structure, so that by systematically varying these orientations the various sterically permissible conformations that were also compatible with the biological, chemical and physical data on the molecule could be identified.

Franklin's insistence on direct analysis of the X-ray data from the fully crystalline A-form and her dismissal of model building, so successful in the discovery of the α -helix, were serious errors of judgement. Crick especially had a strong commitment to model building. It is important to emphasize, in view of the attention that has been given to the exploitation of King's data by the Cambridge workers, that before the discovery of the double helix, Crick encouraged Wilkins and Franklin to build models and provided them with jigs developed in Cambridge for constructing the atomic components.

A particular advantage of molecular models is that they provide a framework within which information from a whole range of physical, chemical and biological techniques can be correlated and integrated. The unfortunate consequences of Franklin's approach in insisting on single-crystal techniques in the analysis of the A form was that she isolated herself from other workers and the information that they could provide. Although Wilkins' trips to Cambridge may well have resulted in a rather one-sided flow of information, they were made in the spirit of the multi-



Part of a letter from Maurice Wilkins to Francis Crick (early 1952) showing a sketch of B-type diffraction from oriented sperm heads. The numbers 1, 2 and 3 identify the first three layer lines. The cross-like distribution of intensity on these layer lines is characteristic of diffraction from a helical structure. On the basis of the angle of the cross, Wilkins estimated the angle of ascent of the helix to be about 40°. Meridional diffraction on higher-order layer lines at the top and bottom of the pattern is also indicated.

disciplinary approach needed for the structure determination of a biological macromolecule such as DNA. The Fourier transform approach to X-ray data analysis from fibres, as suggested by Stokes, also has the advantage over single-crystal methods in that it can reveal at an early stage general features of the molecular structure. It showed in the case of DNA that the molecule is helical, and it yielded parameters defining the helix geometry.

Openness

The history of the discovery of DNA is too often presented in popular accounts in terms of results ‘stolen’ by Watson and Crick, with Franklin as the victim. Yet in the complex interactions in and between the two laboratories, it is not sustainable to view Franklin merely as a victim of other people's actions. When she joined King's she was given the superior material procured by Wilkins from Signer, which had consistently given much better diffraction patterns than any samples



King's reunited: (from left) Raymond Gosling, Herbert Wilson, Maurice Wilkins and Alec Stokes at a celebration for DNA's 40th anniversary.

from other sources. In Gosling she was given a research student who had worked with Wilkins and was already experienced in X-ray data collection and analysis. In Stokes she had access to outstanding expertise in diffraction theory and in Wilkins she had the opportunity to collaborate with someone who already had an international reputation for physical studies on DNA.

It was to the sad detriment not only of herself but also of the King's laboratory as a whole that Franklin chose to work in isolation on a problem, the solution of which depended on confluent results from several workers using different techniques. This is particularly a matter for regret because the experimental work that Franklin performed at King's was of the highest quality; her use of Patterson techniques to obtain structural information from fibre-diffraction data was highly innovative, if disappointing in its outcome. Franklin's approach contrasted markedly with that of Wilkins, who made his results freely known. It is appropriate therefore to end this account with an expression of Wilkins' attitude to science as reflected in his letter to Crick (quoted in ref. 7) after he had visited Cambridge to see the Watson and Crick model.

Dear Francis

I think you're a couple of old rogues but you may well have something. I like the idea. Thanks for the MSS. I was a bit peeved because I was convinced that the 1:1 ratio was significant and had a 4 planar group sketched and was going to look into it and as I was back again on helical schemes I might, given a little time, have got it. But there is *no good grousing* — I think it's a very exciting notion and who the hell got it isn't what matters.

It may well be that historians of science will choose to view this affirmation as a mere epitome of the stiff upper lip. But we must hope that it will be seen for what it is — a principled expression of an attitude, now all too rare, that both the scientific and the wider community should applaud.

Watson Fuller is at the School of Chemistry and Physics, Keele University, Staffordshire ST5 5BG, UK.

1. Watson, J. D. *The Double Helix: A Personal Account of the Discovery of the Structure of DNA* (Athenaeum, New York, 1968).
2. Stokes, A. R. in *DNA Genesis of a Discovery* (ed. Chomet, S.) 27–42 (Newman-Hemisphere, London, 1995). This volume also contains articles of historical interest by M. H. F. Wilkins, R. G. Gosling and H. R. Wilson.
3. Astbury, W. T. *Sym. Soc. Exp. Biol.* **1**, 66–76 (1947).
4. Maddox, B. *Rosalind Franklin — The Dark Lady of DNA* (Harper Collins, London, 2002).
5. Wilson, H. R. *Trends Biochem. Sci.* **13**, 275–278 (1988).
6. Franklin, R. E. & Gosling, R. G. *Acta Cryst.* **6**, 673–677 (1953).
7. Olby, R. *The Path to the Double Helix* (Macmillan, London, 1974).
8. Klug, A. *Nature* **219**, 808–810; 843–844 (1968). See also corrigenda pp. 879 and 1192; correspondence p. 880.
9. Klug, A. *Nature* **248**, 787–788 (1974).
10. Langridge, R. *et al.* *J. Mol. Biol.* **2**, 19–37 (1960).
11. Langridge, R. *et al.* *J. Mol. Biol.* **2**, 38–64 (1960).
12. Fuller, W. *et al.* *J. Mol. Biol.* **12**, 60–80 (1965).

Finding the time

The scientific struggle to bring the world's clocks into line.

Einstein's Clocks and Poincaré's Maps: Empires of Time

by Peter Galison

W. W. Norton: 2003. 370 pp. \$23.95

Sceptre: 2003. £14.99

Jeremy Gray

The challenge of synchronizing clocks around the world, and accurately determining longitudes, required the construction of a great net of cables, often with considerable difficulty. With similar energy, scientists and politicians fought to establish the zero meridian and even the units of time (the French lost both contests, so Greenwich is the zero meridian and time was not decimalized). Peter Galison tells this tale in a tremendous yarn that takes up the largest of the three parts in his new book, *Einstein's Clocks and Poincaré's Maps*. Henri Poincaré was a prominent figure in these tumultuous events, and Galison argues convincingly, backed by a great deal of hitherto neglected information, that he was a leading product of his formative institution, the Ecole Polytechnique in Paris.

But what does this tell us about time? Newton had believed what most people involved in the creation of the global network of clocks believed, namely that the Universe is like a stage on which it makes sense to be able to say precisely when two separate events take place. The great struggle to create a single empire of time in which it was possible to say anywhere in the world what the time was (as recorded on a clock in London) was a move to put Newton's idea of time into operation. Observers in London, Paris and New York could all agree that it was, say, 17:00 (although local times might differ). This process is somewhat arbitrary, but the concept of time underpinning it is fundamentally Newton's, shorn, in Poincaré's view, of some illegitimate intuitions.

On this conception, a fast-moving observer can read off the time by simply looking at standardized clocks as he passes them. But that is not what can be done. Einstein's shattering insight into time is that Newton's universal time does not make sense. Observers in a state of constant relative motion with respect to each other do not keep the same time: their clocks beat at different rates.

By 1900, physicists knew that something was amiss in their theories. Prominent among them was Hendrik Lorentz, who in 1904 introduced a trick to resolve discrepancies in the theory of the moving electron. This was the idea of local time. Poincaré showed in 1905 that Lorentz's local time

could be interpreted as the time kept by an imaginary clock moving with the electron as it beat against the ether. In the same way, lengths of objects were supposed to contract as they moved relative to the ether. By 1912, Poincaré had come close to accepting Einstein's theory, but he regarded it as a convention, to be accepted or rejected on utilitarian grounds, and never gave it the force that the iconoclastic younger man did.

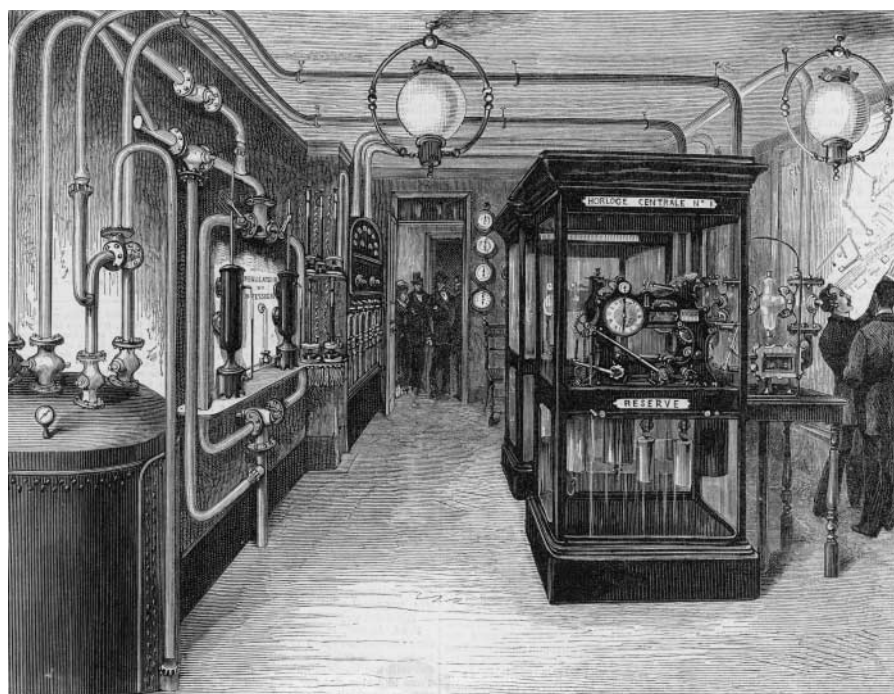
This shift from one conception of time to another is the second theme of the book, and it is less successful. Galison would have us believe that empire time was radically different from newtonian time. He writes: "Newton's absolute, theological time had no place ... Engineering common time stood where God's absolute time had been," but this seems unconvincing. Galison also downplays the details of the intense scientific debate, which was well described in Arthur I. Miller's *Albert Einstein's Special Theory of Relativity* (Addison-Wesley, 1980), in order to give salience to the technological and political factors he has identified. In so doing he pushes Poincaré and Einstein closer together than they were, stressing their involvement in the world of clocks and measurements, and diminishing the radical nature of Einstein's ideas.

Einstein is presented here as a serious and talented worker in the patent office. This is surely an improvement on his portrayal, sometimes fostered by Einstein himself, as

an other-worldly academic, although Lewis Pyenson has already described Einstein as an enthusiastic worker in the patent office in *The Young Einstein* (Adam Hilger, 1985). Perhaps Einstein's daily working life contributed to his insistence on clocks and the measurement of time in his famous paper of 1905 in which he introduced special relativity. However, he is also described as a radical and an iconoclast. The inevitable question is then to decide what weight to attach to these different factors, and Galison refuses to address the matter.

Instead, he has written a methodological final chapter arguing for the inseparability of all the factors that he has introduced, from the grittiest to the most ethereal. He argues that Poincaré in particular was situated at the intersection of three arcs, namely developments in physics, telegraphy and philosophy. Einstein was likewise at a triple intersection, of a slightly different kind. We can accept that it is not easy to situate scientific discovery in a richly overlapping set of contexts, and that conclusions should be tentative, but by marginalizing the strictly scientific debate, the implication that the other arcs carry significant weight is allowed to slip through too easily. This matters greatly when discussing Poincaré, because it is often argued that his capacity to move between mathematics, physics and philosophy often weakened his insights into physics.

The logic of Galison's position is surely



Keeping time: the control room at the Rue du Télégraphe in Paris synchronized the city's clocks.

MARY EVANS

that, if empire time is by its nature different from newtonian time, then one might expect someone heavily committed to the establishment of empire time to differ from those habituated to newtonian time. Poincaré should differ from Lorentz, at least, and be closer to Einstein. If empire time was simply understood as being newtonian time put into practice, then Poincaré would resemble Lorentz and differ from Einstein, who believed that time had to be defined by means of clocks (and not merely captured).

Perhaps Poincaré's deep intellectual and political immersion in the creation of empire time with its global sense of simultaneity worked against him when he had the chance to reformulate time as Einstein did, and even to accept it afterwards as, steadily, his contemporaries did. But there are other possible explanations and such inferences are risky. Ultimately, the methodological sophistication in this book may not deliver what its author hopes, but the wealth of information it contains is surely stimulating. ■

Jeremy Gray is at the Centre for the History of the Mathematical Sciences, Faculty of Mathematics and Computing, Open University, Milton Keynes MK7 6AA, UK.

Fountain of hype

Merchants of Immortality: Chasing the Dream of Human Life Extension

by Stephen Hall

Houghton Mifflin: 2003. 440 pp. \$25

S. Jay Olshansky

One can imagine an alchemist some 2,000 years ago weaving a tale for the investors and politicians of his day. They cling to every word as he tells them how the properties of immortality thought to exist for gold could be transferred to humans by ingesting minute quantities of the rare metal. According to the alchemist, 'science' is only 10–20 years away from a breakthrough that would give rise to immortality for those willing to invest. All that is needed is money and time to figure out the recipe.

Such merchants of immortality have been spinning their tale since time immemorial. So it is not surprising that in today's rapidly ageing society, where medical technology can manufacture survival time, there is an abundance of scientists and longevity salesmen (on occasion they are one and the same) who seek lucre from the promise of a longer and healthier life. Are the modern immortality salesmen on the trail of the genuine fountain of youth, or are contemporary investors and consumers of the anti-ageing industry being duped, like countless before them? In *Merchants of Immortality*, Stephen Hall provides the answer.



Making a splash: immortalists make dramatic claims that there really is a fountain of youth.

Hall begins with an account of Leonard Hayflick's discovery that some cells have a limited capacity to replicate. The irony is that Hayflick, who is the father of biogerontology, is also the most ardent voice of authority against the immortalists, yet his view on the topic is barely mentioned. Instead Hall goes into detail about Hayflick's battle with the government over a human cell line he developed that was later used to produce life-saving vaccines.

The one scientist who appears repeatedly throughout the book as the most ardent merchant of immortality is Michael West, the budding priest turned scientist and entrepreneur who wrote about the longevity-enhancing effects of telomerase, cloning and embryonic stem cells. Hall takes issue with West's recipe for immortality, which has lured in a bevy of investors who were all too willing to have a financial stake in the fountain of youth.

Several chapters are devoted to similar tales of scientists-turned-entrepreneurs. They are mainly fascinating but sad stories of how money both corrupts and enhances modern science. But it is exactly this kind of entrepreneurship that, despite its potential to corrupt and the numerous roadblocks placed before science by politics and ignorance, has the potential to create a new field of regenerative medicine from recent developments regarding embryonic and adult stem cells.

It is often difficult for scientists to see the importance of their work and its relevance in a world where social norms and ethics

can mesh and clash at the same time. There is no area of research where this dichotomy is more evident than studies of ageing. Hall pays what seems at first to be an inordinate amount of attention to the politics behind cloning and the use of embryonic and adult stem cells, but it soon becomes evident how debates about abortion and right-to-life issues have a direct influence on funding for research. He stresses that the politics of science in this case is not about the quest for immortality or the battle against ageing itself, but about efforts to combat disease.

Hall makes the case that the modern search for the fountain of youth, through the study of telomeres and embryonic and adult stem cells, is built on a house of cards. It is a mirage, he says, offering everlasting (or at least much longer) life, and has been placed on the horizon by a handful of scientists who seek wealth by making some exaggerated claims. What may realistically be on offer is not an extended lifespan but rather a healthier old age. Hall mentions scientists (including Hayflick) who have challenged the remarkable claims of the immortalists. But given the tales being spun and the pot of money awaiting a victor in the quest for a fountain of youth, it is no surprise that he has chosen to focus on the immortalists.

The field of ageing research should be grateful to Hall for applying his journalistic wizardry to the story of the modern quest for immortality. In this book he captures the drama, excitement, competition and

rewards, both personal and professional, of scientific research. He treats the reader to a behind-the-scenes look at the jealousy, backstabbing, ambition and embellishment that comes with the territory of the modern scientific quest to halt or turn back the hands of time. In the end, Hall makes it clear that although today's immortalists may be whispering optimistically about their own immortality as they take their own last breath, there is a glimmer of hope that the true fountain of youth could be just over the horizon.

Merchants of Immortality is a remarkable book that is a must-read in the world of science and medicine. And lay readers will be interested in this revealing story about the odd mixture of science, politics and ethics, and how they all come together in the modern quest for immortality. ■

S. Jay Olshansky is at the School of Public Health, University of Illinois at Chicago, Chicago, Illinois, USA.

Ecology you can count on

Population Ecology: First Principles

by John H. Vandermeer & Deborah E. Goldberg

Princeton University Press: 2003. 296 pp. \$75, £52 (hbk); \$35, £24.95 (pbk)

Alan Hastings

A mathematical and conceptual approach to population ecology — the study of the dynamics of animals and plants in nature — can be traced back to work early in the twentieth century. The fundamentals were developed by a number of scientists: Alfred J. Lotka, Vito Volterra and G. F. Gause worked on the dynamics of single and interacting populations; A. J. Nicholson and V. A. Bailey studied host-parasitoid dynamics; Lotka and P. H. Leslie did the early work on age-structured populations; J. G. Skellam did the same for spatial dynamics; and W. O. Kermack and A. G. McKendrick, and then Norman T. J. Bailey and others, worked on the dynamics of diseases and epidemics.

These innovators set the stage for recent advances in the understanding of the dynamics of ecological populations, and in applied areas such as the management of renewable resources, conservation biology and understanding the population dynamics of diseases. Progress in all these areas has come from a tight coupling of mathematical theory with biology, and students and ecologists need a good grounding in these areas. How well does the book *Population Ecology* succeed in developing the field from a set

of first principles, and who is most likely to profit from reading it?

Research based on first principles in population ecology depends, at the very least, on a thorough understanding of basic calculus and linear algebra, and arguably needs more advanced techniques as well, such as stochastic processes and partial differential equations. So it is important how the book treats these areas, and what knowledge it assumes on the part of the reader. The stated assumption is that readers are familiar with calculus but may not have seen the required results in linear algebra — there is a brief appendix summarizing the major results for linear algebra, but not for calculus. I think that to follow the mathematical development, readers will need a good to outstanding working knowledge of calculus and be comfortable with mathematical manipulations.

In some respects, the chapter on dynamics emerges as one of the strongest parts of the book; in others, it is the weakest. It is a truly beautiful exposition of nonlinear dynamics at a relatively elementary level, but unfortunately it is almost devoid of biology. And in population ecology, where mathematics and biology are intimately connected, it is important that these disciplines are well integrated. The rest of the book does this much better, helping readers to understand how both need to be used to get at the heart of population ecology. The chapter on space and metapopulations goes further, by explaining the importance of a statistical approach to understanding spatial pattern, and in so doing it emphasizes the role of model formulation.

As is typical in ecology, many mathematical arguments are presented graphically. Some of these are well done, for example in the chapter on competition models, but some are confusing, such as those illustrating the components of predator-prey interactions. In a few places, such as the chapter on life histories, a more graphical development would have been welcomed.

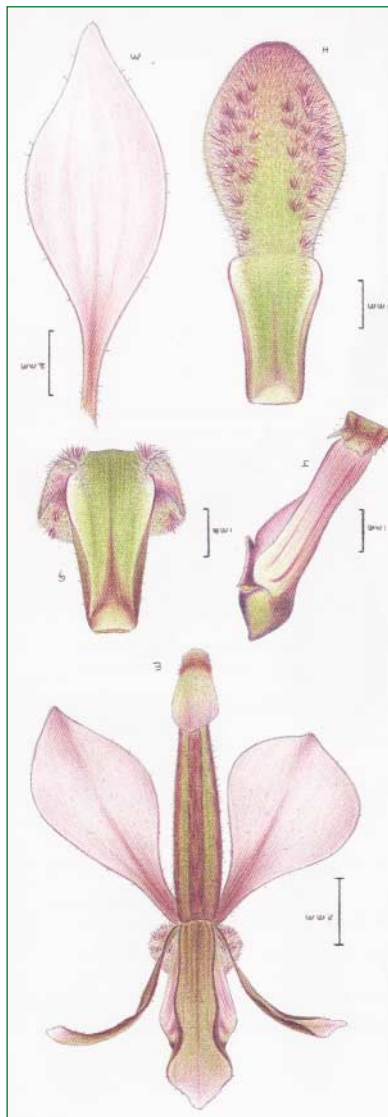
Overall, the book provides a solid foundation for further study in population ecology, and provides good coverage of all the fundamental areas of population ecology outlined above. But as a textbook, or for individual study, there is room for improvement. The lack of problems to solve is a big shortcoming, as mathematical developments are best understood by doing, rather than by merely reading. And readers who are not advanced mathematically might find several places where the mathematical development is too terse, for example in the chapter on projection matrices. Also, the book doesn't use many different biological examples to bolster an argument, just a few choice studies (very few in the chapter on competition). Yet the overall concept of emphasizing first principles in population ecology is very appealing, and readers who

understand the material presented will be able to cope with most current research in population ecology. ■

Alan Hastings is in the Department of Environmental Science and Policy, University of California, Davis, California 95616, USA. He is the author of *Population Biology: Concepts and Models*.

Petal power

Illustrations of orchids, such as those of *Eriochilus cucullatus* shown here, are normally found in historic texts predating modern photography. But Princeton University Press has just published the first in a series of books on the *Orchids of Australia*, each containing new illustrations of some 150 species by artist John J. Riley. Accompanied by descriptive text by Riley and David P. Banks, these books are intended to provide a visual diagnostic reference for orchid enthusiasts, as well as a feast for admirers of botanical illustration as art.



Understanding cells

Paul Nurse

Many of the properties that characterize living organisms are also exhibited by individual cells. These include communication, homeostasis, spatial and temporal organization, reproduction, and adaptation to external stimuli. Biological explanations of these complex phenomena are often based on the logical and informational processes that underpin the mechanisms involved. Two examples of this are the significance of the structure of DNA, and the mechanisms that control gene expression. DNA's structure relates to heredity because of the coding and replicative capacity of its polynucleotide sequence, whereas the interactions of activators and repressors with promoter regions are best understood in terms of the feedback loops that regulate gene expression.

Most experimental investigations of cells, however, do not readily yield such explanations, because they usually put greater emphasis on molecular and biochemical descriptions of phenomena. To explain logical and informational processes on a cellular level, therefore, we need to devise new ways to obtain and analyse data, particularly those generated by genomic and post-genomic studies.

An important part of the search for such explanations is the identification, characterization and classification of the logical and informational modules that operate in cells. For example, the types of modules that may be involved in the dynamics of intracellular communication include feedback loops,

switches, timers, oscillators and amplifiers. Many of these could be similar in formal structure to those already studied in the development of machine theory, computing and electronic circuitry. When these modules are coupled in space by processes such as reaction diffusion and regulated cytoskeletal transport, they help to provide a basis for the spatial organization of the cell. The identification and characterization of these modules will require extensive experimental investigation, followed by realistic modelling of the processes involved. Such analyses would allow a catalogue of the module types that operate in cells to be assembled.

The next task is to identify the interacting molecules and biochemical activities that generate the characteristic behaviours of particular modules. This will be more difficult, involving systematic mapping of particular module types to specific molecules, biochemical activities and molecular interactions, and assembling the information into databases. Some examples may be useful in thinking about this. Proteins that can be attacked by proteases are likely to be found in switching modules, as destruction of a protein can bring about an irreversible switch in the behaviour of a regulatory network. Small G-protein GTPases that bind to and then convert GTP to GDP could act as timers because of the time it takes to execute this conversion. In turn, GTPase timers can act as amplifiers if signals are continually sent when in the GTP-bound state, or as components of proofreading modules, as seen during protein translation. Antagonistically acting protein kinases and phosphatases could also act as amplifiers and switches.

If sufficient regularity can be found between molecular entities and logical and informational outcomes to allow appropriate databases to be built, then genomic and post-genomic data could be interrogated more effectively. Generally, three types of information are found in genomic and post-genomic databases: sequence data that define molecules and biochemical activities; interaction data based on co-localization, co-expression and analyses of interacting pairs of proteins (two-hybrid analyses); and functional data based on gene deletions and RNA-interference studies. These data allow specific molecules to be linked to particular biological phenomena.

The objective is to use this information to identify the molecular assemblies involved in a biological phenomenon of interest, and then to determine the likelihood of that assembly being part of a particular logical and informational module. If successful, this approach would not require detailed kinetic

Systems biology

New approaches are needed to determine the logical and informational processes that underpin cellular behaviour.

analyses of all processes within cells, but rather would rely on more cursory calculations to study the phenomena of interest. The rough nature of the calculations would be justified by comparison with well-understood examples of particular modules.

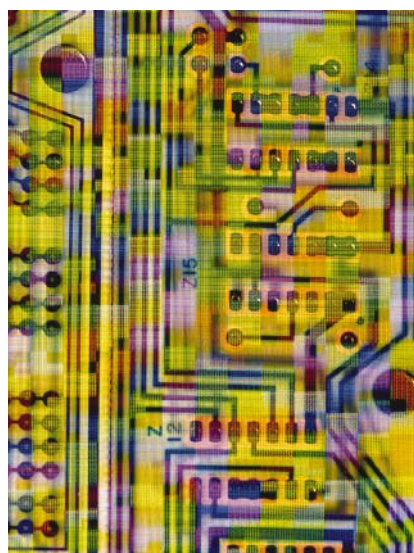
A useful analogy of what is being proposed is the analysis of an electronic circuit. Once the detailed operations of different types of electronic components have been identified, it is possible to gain insight into what an electronic circuit can do simply by knowing what components are present and how they are connected, even if their precise dynamic behaviour has not been determined. The various logical and informational modules implicated in a biological phenomenon of interest have to be integrated in order to generate a better understanding of how cells work. One process that has been analysed comprehensively in this way is bacterial chemotaxis.

The success of this general approach depends on there being a limited set of biochemical activities and molecular interactions that together can solve the myriad logical and informational problems found in biological systems. If there is only a restricted set of processes that are efficient and stable in operation and which have been exploited by evolution, then there should be only a limited set of possible solutions to real biological problems. Of course, if nature shows no such restraint, then we must go back to the drawing-board if we are ever to understand its complexity.

Paul Nurse is at the Cell Cycle Laboratory, Cancer Research UK, Lincoln's Inn Fields, London WC2A 3PX, UK.

FURTHER READING

- Bourne, H. R., Sanders, D. A. & McCormick, F. *Nature* **348**, 125–132 (1990).
 Bray, D. *Proc. Natl Acad. Sci. USA* **99**, 7–9 (2002).
 Guet, C. C., Elowitz, M. B., Hsing, W. & Leibler, S. *Science* **296**, 1466–1470 (2002).
 Kauffman, S. A. *The Origins of Order: Self-Organization and Selection in Evolution* (Oxford Univ. Press, 1993).
 Novák, B. & Tyson, J. J. *J. Cell Sci.* **106**, 1153–1168 (1993).
 Tyson, J. J., Chen, K. C. & Novak, B. *Curr. Opin. Cell Biol.* **5**, 221–231 (2003).



Cellular processes may be studied in a similar way to analysing an electronic circuit.

T. CRADDOCK/SPFL

You aren't what you eat

Henry Gee

An obscure marine worm does not belong among the molluscs, as had been thought. Rather, it has a claim to being the most primitive extant member of the group of animals that includes vertebrates.

Only those fond of the less-frequented pages of zoology textbooks — the chapters that deal with organisms such as gastrotrichs and gnathostomulids — will have ever heard of a small marine worm called *Xenoturbella bocki*. First described in 1949, this creature has struggled to find a phylogenetic home, and has been loosely attached to many animal groups. The problem is its unassuming nature. *Sans* proper gonads, excretory system, body cavity, even a through gut, the worm has very few distinguishing features. And it is only about 2 cm in length. In the end, zoologists usually admitted defeat and banished it to one of several groups of primitive flatworm.

In 1997, however, it was announced that *Xenoturbella* is, in fact, a mollusc, and not only that, a bivalve mollusc^{1,2}, even if, like the persona of Eliza Doolittle in *My Fair Lady*, neither Hungarian nor of royal blood. DNA from the creature closely resembled that of the small bivalve *Nucula*, and — this was the clincher — mollusc-like eggs were found inside *Xenoturbella* specimens. It was thought that the animal started life as a conventional mollusc, but that by the time it reached adulthood it had for some reason shed virtually everything that would mark it as such.

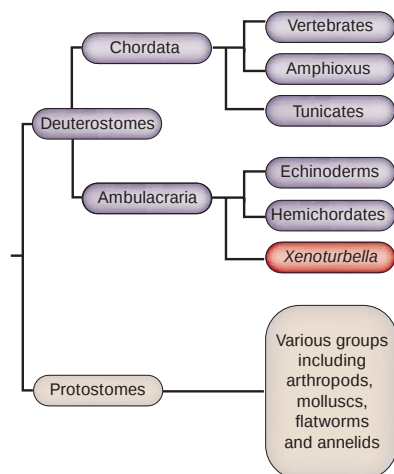


Figure 2 A place for *Xenoturbella* among the deuterostomes. Bourlat *et al.*³ place this worm-like creature as an early offshoot of the ambulacraria. In previous thinking it had been found a home among the flatworms or molluscs.



Figure 1 Spot the ordinary flower-girl. Audrey Hepburn as Eliza Doolittle, and *Xenoturbella*.



DNA, like the camera, cannot lie: scientists, on the other hand, are entitled to change their minds. A reappraisal of *Xenoturbella* DNA by Bourlat and colleagues (page 925 of this issue³) shows that if it contains mollusc-like DNA and eggs, this is because it feeds on molluscs, not because it is a mollusc itself. *Nucula* has DNA that is virtually identical to that found in *Xenoturbella*, and is abundant in those places where the worm is found. Once mollusc DNA is accounted for, a residue of genuine *Xenoturbella* DNA can be studied, and the results are even more surprising than the supposed parentage of Eliza Doolittle: *Xenoturbella* is no ordinary flower-girl (Fig. 1), but arguably the most primitive extant member of the deuterostomes. This is the great group of animals to which we ourselves belong, the other major group being the protostomes (which contains arthropods and annelids).

The deuterostomes are themselves divided into two major lineages (Fig. 2). One leads to the chordates, which include tunicates, the fish-like amphioxus and vertebrates, including humans. The other branch, known as the ambulacraria, contains echinoderms (starfishes and allies) and the hemichordates, which consist of an assortment of colonial forms called pterobranchs, and free-living marine worms known as enteropneusts. Using a variety of arguments based on DNA sequence comparison, gene order and codon usage, Bourlat and colleagues place *Xenoturbella* as an early offshoot of the ambulacraria, before it diverged into echinoderms and hemichordates (Fig. 2).

This assignment makes the anatomy

(such as it is) and molecular biology of *Xenoturbella* of intense interest to evolutionary biologists investigating the origin of vertebrates and the early history of deuterostomes. These have been extraordinarily hard problems to crack. The deuterostomes alive today reveal little about the group's past diversity, which may have been much greater than we see now. The best evidence of this comes from the echinoderms, which, because of their calcite skeletons, have left a rich fossil record. There are only five or six extant classes of echinoderm; in the Ordovician, around half a billion years ago, there were 20 or more. But many deuterostomes are soft-bodied. And those few fossils that exist are hard to interpret because of a tendency in deuterostomes towards radical transformation of body form, especially in adults. There is no reason to suppose that early deuterostomes were not as divergent as extant ones, as for example evidenced by the vetulicolians — strange, ancient fossils claimed to be early deuterostomes⁴. This divergence, combined with the paucity of extant forms that means we lack modern models on which to base anatomical analyses of fossils, makes interpreting ancient deuterostome fossils difficult. Having an extant, primitive example to study could therefore be of immense value in tracing the evolution of deuterostome groups.

However, the simplicity of form in *Xenoturbella* will perhaps be of more use in trying to understand what the earliest common ancestor of deuterostomes looked like. Deuterostomes and protostomes have a common heritage of genes, such as the *Hox*

group, that are concerned with segmental identity, leading to a view of evolution in which most animals evolved from a segmented animal with a relatively complex anatomy. This view has been challenged, however, not least by work on enteropneusts showing that anatomy is no guide to the complexity of gene expression⁵. *Xenoturbella* could be a model of early deuterostome ancestry, if not the ancestry of both deuterostomes and protostomes. Classifying it as a flatworm, therefore, might not have been so far from the truth, as that is what the earliest deuterostomes may have looked like.

However, the fact that Bourlat and colleagues³ have placed *Xenoturbella* closer to the ambulacraria than to the root of deuterostomes in general raises further problems, for it means that it cannot represent a truly primitive state. Ambulacraria and chordates share anatomical features, such as a body cavity and structures called pharyngeal slits,

that were presumably primitively present in the *Xenoturbella* lineage but have since been lost. This means that *Xenoturbella* is secondarily simplified rather than pristine and primitive, and may not be as reliable a guide to deuterostome evolution as we should like.

As we reflect on *Xenoturbella*'s newly exalted state as a deuterostome, it is worth recalling similar examples. Aristotle classed tunicates with the molluscs, and the amphioxus, when first described, was named *Limax lanceolatus* — a kind of slug. So there is no shame in having once been classed as a mollusc: many of the best deuterostomes started out that way. ■

Henry Gee is a senior editor at Nature.

1. Norén, M. & Jondelius, U. *Nature* **390**, 31–32 (1997).
2. Israelsson, O. *Nature* **390**, 32 (1997).
3. Bourlat, S. J., Nielsen, C., Lockyer, A. E., Littlewood, D. T. J. & Telford, M. J. *Nature* **424**, 925–928 (2003).
4. Shu, D.-G. *et al.* *Nature* **414**, 419–424 (2001).
5. Lowe, C. J. *et al.* *Cell* **113**, 853–865 (2003).

Granular materials

Shaken sand — a granular fluid?

Paul Umbanhowar

The connection between random grain motion and viscosity in shaken sand — a strongly non-equilibrium system — has been probed. Curiously, the link is similar to that found in an ordinary liquid in thermal equilibrium.

By measuring both the free and forced oscillations of a rigid pendulum immersed in an ordinary liquid, the temperature and viscosity of the liquid can be determined¹. This is due, in part, to a relation from equilibrium statistical mechanics known as the fluctuation–dissipation theorem², which, in a precursor to its modern form, was devised by Einstein³ to explain the diffusive Brownian motion of small particles suspended in liquids⁴. Driven granular materials, such as shaken sand, are systems far from equilibrium — they have strong spatial and temporal variations in quantities such as density and local particle velocity, and would consequently not be expected to obey the fluctuation–dissipation theorem.

Yet, in many instances, the macroscopic behaviour of granular materials seems, at least superficially, liquid-like⁵. One might then wonder whether an experiment using a rigid pendulum would reveal a similar fluctuation–dissipation relation in driven granular materials, despite their dissipative nature. On page 909 of this issue⁶, D'Anna and colleagues take up this question and find, surprisingly, that the answer seems to be yes. In particular, their experiments show that the free and forced motions of the probe are related by a fluctuation–dissipation-like relation, and that an effective viscosity and temperature can be defined.

Granular materials — such as peas, coal,

pillars, breakfast cereal and, not least of all, sand — are usually defined to be discrete solid bits that interact with each other through energy-dissipating contact forces (although the definition is sometimes stretched to include wetted grains and powders for which attractive surface forces are also important). Many industrial practices require the efficient handling and mixing of granular materials: food and agricultural processing, sorting and assembly of parts, cement manufacturing, mining, radiation shielding and, of ever increasing scope, pharmaceutical production. Despite their somewhat humble nature, granular materials behave unusually because they combine properties of both liquids and solids^{7,8}. Examples of such behaviour include the ability to de-mix when poured⁹, to form waves or ripples when shaken¹⁰ or blown¹¹, and to expand when squeezed¹² (as anyone can attest who has walked on wet sand and observed a dry halo around their foot).

Many of the unique properties of granular materials arise because, unlike an ordinary fluid, the kinetic energy associated with the relative motion of macroscopic particles (called 'granular temperature' by physicists) is not constant. Instead, it is continually and irrevocably transferred by collisions to internal (thermal or non-kinetic) degrees of freedom. Although individual grains possess a well-defined thermodynamic temperature,

the associated thermal energy (equal to the product kT , where k is Boltzmann's constant and T is the temperature in kelvin) is too small to allow relative particle motion, as it does in ordinary liquids. So dynamic collections of grains require a continuous external source of energy to prevent them from getting stuck in a particular configuration.

To complicate matters further, because energy is typically injected through a surface (a stirring rod, shear zone or vibrating wall, for example), steep gradients of kinetic energy are invariably present in the system, caused by the decrease in relative particle motion further from the energy source and by the spatial anisotropies arising from the directional nature of the forcing. These complications make it unclear whether granular liquids can be described by the traditional hydrodynamics derived from equilibrium statistical mechanics^{5,7,8}. And they are the major reasons why granular materials remain one of the least well understood classes of matter.

Rather than being deterred by the complications associated with the lack of energy conservation in granular media, D'Anna *et al.*⁶ have attacked the problem directly by studying the response of vertically vibrated glass beads (physicists' equivalent of sand), using a beefed-up version of the aforementioned pendulum (pictured on the cover of this issue). Their bead-filled container is driven with band-filtered white noise at a relatively high frequency, which ensures that no single characteristic frequency dominates the forcing, that the natural frequency of the oscillator is far below the driving frequency, and that the beads are condensed. The pendulum probe is operated in two modes: free and forced. In the free mode, a constant barrage of grains bombards the probe and jostles it about in irregular excursions about its equilibrium position. Analysis of this Brownian-like motion gives the 'noise power spectral density' in terms of the driving frequency. In the forced mode, a sinusoidal torque is applied to the oscillator, and measurements of the relative phase and amplitude of the response determine the 'complex susceptibility'.

Interestingly, the group finds that the measured complex susceptibility is well described by the theory developed for equilibrium fluids — one of the two parameters used in the fit to theory being the friction coefficient, which is proportional to the shaken beads' viscosity. Furthermore, the bead viscosity is found to decrease nearly linearly with increasing amplitude of shaking. Ordinary liquids show a similar decrease in viscosity with increasing temperature — nearly a factor of three for water between 0°C and 45°C.

Of greater interest is D'Anna and colleagues' finding⁶ concerning the so-called fluctuation–dissipation ratio, which is

derived from the complex susceptibility and the noise power spectral density, and so relates collective response to individual grain motion. This ratio turns out to be nearly constant for varying frequency, which enables D'Anna *et al.* to define an effective temperature for the grains. For a classical liquid, this ratio is equal to kT and is strictly independent of frequency. Using the average value of the ratio as their measure of thermal energy, the authors show that the temperature increases as the square of the driving-force amplitude — a result that would be naively expected if the bead velocities were linearly related to the container velocity, which in turn increases in direct proportion to the drive amplitude. These findings are intriguing, and they support results from other analyses of model systems that have indicated that the fluctuation–dissipation theorem, or a slightly modified version of it, applies to granular materials^{13–15}. There are also signs that the temperatures obtained through application of the fluctuation–dissipation theorem¹³ are compatible with the temperature obtained from a new form of statistical mechanics that is applicable to granular systems and possibly to other non-energy-conserving systems¹⁶.

D'Anna and colleagues' findings that granular ensembles with strong dissipation have a definable viscosity and approximately obey the fluctuation–dissipation theorem are exciting, but they do not yet definitively answer the question of how deep the similarities run between moving grains and ordinary

liquids. Some puzzles remain. Why does the effective temperature vary by approximately a factor of 10 for differently shaped probes? What influence does the probe have on the measurements? Why is the fluctuation–dissipation ratio an increasing function of frequency (albeit slowly)? The answers themselves may be mundane, but they might lead to deeper insights into the properties of granular materials and other related non-equilibrium subjects such as traffic flow, flocking, evolving networks and turbulence¹⁴.

Paul Umbanhowar is in the Department of Physics and Astronomy, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208-3112, USA.

e-mail: umbanhowar@nwu.edu

- Uhlenbeck, G. E. & Goudsmit, S. *Phys. Rev.* **34**, 145–151 (1929).
- Reichl, L. E. *A Modern Course in Statistical Physics* 545–560 (Univ. Texas Press, Austin, 1991).
- Einstein, A. *Ann. Physik* **17**, 549–560 (1905).
- Brown, R. *Phil. Mag.* **4**, 161–173 (1828).
- Jaeger, H. M., Nagel, S. R. & Behringer, R. P. *Rev. Mod. Phys.* **68**, 1259–1273 (1996).
- D'Anna, G., Mayor, P., Barrat, A., Loreto, V. & Nori, F. *Nature* **424**, 909–912 (2003).
- Kadanoff, L. P. *Rev. Mod. Phys.* **71**, 435–444 (1999).
- de Gennes, P. G. *Rev. Mod. Phys.* **71**, S374–S382 (1999).
- Makse, H. A., Havlin, S., King, P. R. & Stanley, H. E. *Nature* **386**, 379–382 (1997).
- Melo, F., Umbanhowar, P. B. & Swinney, H. L. *Phys. Rev. Lett.* **75**, 3838–3841 (1995).
- Bagnold, R. A. *The Physics of Blown Sand and Desert Dunes* (Methuen, London, 1941).
- Reynolds, O. *Phil. Mag.* **20**, 469 (1885).
- Makse, H. A. & Kurchan, J. *Nature* **415**, 614–617 (2002).
- Aumaitre, S., Fauve, S., McNamara, S. & Poggi, P. *Eur. Phys. J. B* **19**, 449–460 (2001).
- Colizza, V., Barrat, A. & Loreto, V. *Phys. Rev. E* **65**, 050301(R) (2002).
- Edwards, S. F. in *Granular Matter: An Interdisciplinary Approach* (ed. Mehta, A.) 121–140 (Springer, New York, 1994).

the two drugs most commonly used to treat malaria⁴. Hope for the containment of the disease now rests largely on a remarkable set of artemisinin drugs developed by Chinese scientists in the 1970s and early 1980s, which rapidly kill the malaria parasites.

The parasites are small 'protozoan' cells (the most prevalent species infecting humans are *Plasmodium falciparum* and *P. vivax*), which enter their human host through a mosquito bite. They first invade the liver and replicate there for two weeks, before beginning a cycle of red-blood-cell invasion, then growth, replication and red-cell destruction that leads to the disease symptoms. The artemisinin drugs are known to act specifically during this blood stage.

Artemisinin contains a structural feature called a peroxide bridge (Fig. 1), and this is believed to be the key to the drug's mode of action. Ferrous iron (Fe^{2+}) catalyses the cleavage of this bridge, forming highly reactive free radicals⁵. The theory has been that these artemisinin-derived free radicals chemically modify and inhibit a variety of parasite molecules, resulting in parasite death^{5,6}.

A rich source of intracellular Fe^{2+} is haem — an essential component of haemoglobin — and it has long been suspected that Fe^{2+} -haem is responsible for activating artemisinins inside the parasite. In support of this, Fe^{2+} -haem activates artemisinins in the test tube and haem–artemisinin complexes can be formed. This theory appealed to malarialogists because it seemed to explain the specificity of the drug within the context of a unique aspect of parasite metabolism. During its growth and replication inside the red blood cell, the parasite ingests and degrades up to 80% of host-cell haemoglobin in a compartment called a food vacuole. This releases Fe^{2+} -haem, which is oxidized to Fe^{3+} -haematin and then aggregates within the food vacuole into an ordered crystalline pigment called haemozoin. A theory developed that the specific antimalarial effect of artemisinin was due to its entry into the parasite food vacuole and its interaction with Fe^{2+} -haem. Here, it would set off a 'cluster bomb' of free radicals, inhibiting several key parasite components and eventually resulting in parasite death.

This theory has been challenged⁷, however,

Malaria

To kill a parasite

Robert G. Ridley

Artemisinins have been used since ancient times to treat malaria. A new theory could explain how this age-old medicine is able to cause the death of the malaria parasite.

The Chinese herb qinghao (*Artemisia annua*) has long been used to treat malaria — Taoist manuscripts dating back to the third century describe the use of qinghao extracts to treat malaria-related fevers¹. Over the past two decades, derivatives of the herb's active ingredient, artemisinin, have made an increasing contribution to malaria treatment. But the precise mechanism by which artemisinin derivatives kill the parasite has remained obscure. Writing on page 957 of this issue, Krishna and colleagues² propose a radical new theory to explain the molecular basis of the antimalarial activity of artemisinin.

Malaria remains a scourge of the developing world, killing over a million people each year and infecting around 500 million³. Most of the victims are children under the age of

five living in sub-Saharan Africa, but the disease also afflicts Southeast Asia, South America and the Indian subcontinent. The situation has worsened over recent years as resistance has developed against chloroquine and sulphadoxine-pyrimethamine,

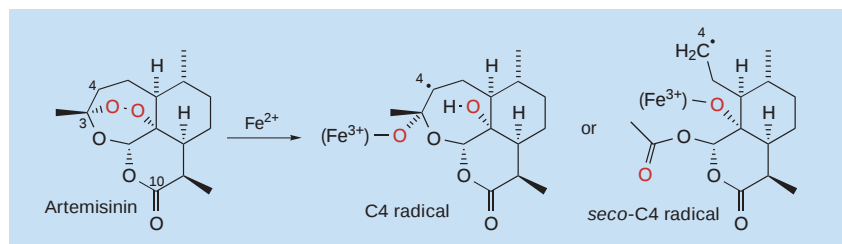


Figure 1 Structure of artemisinin. The molecule contains a peroxide bridge (red), which becomes cleaved when artemisinin interacts with ferrous iron (Fe^{2+}). Cleavage creates 'C4' and 'seco-C4' free radicals, each capable of chemically modifying biological molecules.

based on chemical structure–activity considerations. The work of Krishna and colleagues² on *P. falciparum* now prompts a further rethink. Using a fluorescently labelled artemisinin derivative and powerful high-resolution confocal microscopy, the authors show that the artemisinins do not accumulate in the food vacuole, but are instead spread throughout the parasite. They further show that inhibitors that prevent haemoglobin degradation and the consequent release of haem do not interfere with the action of artemisinins. These inhibitors do antagonize the antimalarial properties of chloroquine, which is known to exert its effect through haem binding⁴. The authors' conclusion supports that of others⁷, that the activity of the artemisinins does not require haem.

If the artemisinins neither interact with haem nor localize in the food vacuole, how do they function? Krishna and colleagues present compelling evidence that the artemisinins work through irreversibly inhibiting a metabolic enzyme — the malarial calcium-dependent ATPase (PfATP6). PfATP6 is very similar to a mammalian ATPase (the sarco/endoplasmic reticulum calcium-dependent ATPase, SERCA), which, as its name suggests, is located in a membrane-enclosed intracellular compartment called the sarco/endoplasmic reticulum (ER). In the parasite, the ER is situated outside the food vacuole (throughout the parasite cytoplasm) — the same location as the proposed site of artemisinin action.

But another connection led the authors to propose that PfATP6 might be a molecular target for artemisinin. The mammalian enzyme SERCA is inhibited by a drug called thapsigargin that, although it lacks a peroxide bridge, has some structural similarities to the artemisinins. To test whether PfATP6 could be a molecular target for artemisinins, Krishna and colleagues introduced the malarial enzyme into frog eggs and assessed its activity in the presence of thapsigargin or artemisinin. Both drugs did indeed inhibit PfATP6. Importantly, thapsigargin also interfered with the action of artemisinin, indicating that both drugs were operating on malaria parasites through similar mechanisms. An iron chelator, desferrioxamine (which removes free iron from the cellular environment), also interfered with artemisinin action — it antagonized the inhibition of PfATP6, the inhibition of parasite growth and the accumulation of artemisinin inside the parasite. Taken together, these data indicate that an iron-dependent mechanism generates free radicals from artemisinin, which inhibit PfATP6, thereby slowing parasite growth (Fig. 2).

Krishna and colleagues make a strong case for their new hypothesis, but certain questions remain and some confirmatory experiments suggest themselves. We still do

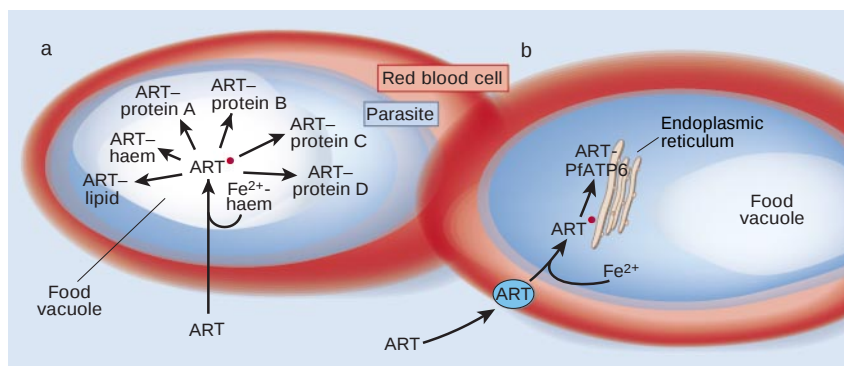


Figure 2 A new molecular target for an ancient antimalarial drug. These images represent **a** (trophozoite-stage) malarial parasite (blue) growing inside a red blood cell. **a**, It was originally supposed that artemisinin (ART) was transported to the food vacuole of the parasite (white), where it was converted into a free radical after an interaction with Fe^{2+} -haem. These free radicals were then thought to modify and inhibit haem, lipids and at least four proteins, resulting in parasite death. **b**, Krishna and colleagues² now propose that artemisinin is transported from the red blood cell into the parasite inside parasite-derived membrane vesicles. Once inside the parasite, artemisinin is activated by free iron, or another iron-dependent process, that occurs close to PfATP6 in the endoplasmic reticulum. The activated artemisinin specifically and irreversibly binds to and inhibits PfATP6, and inhibits parasite growth.

not have definitive biochemical evidence that the activated artemisinins bind to PfATP6 inside the parasite at inhibitory concentrations. We also do not know whether irreversible binding occurs at one or more sites. The fact that resistance has not yet developed to the artemisinins suggests that their binding to PfATP6 is not affected by single point mutations in the target protein. So the free radicals generated from artemisinin might indeed modify multiple sites on a single target. Furthermore, other studies have shown that artemisinin can bind to several different parasite proteins⁵, so it is possible that the drug has other molecular targets, in addition to PfATP6. In short, further biochemical work with parasite material is needed.

Nevertheless, this study² will radically alter our thinking about the mode of action of artemisinins. PfATP6 is also now a prospective target for the development of

new antimalarial drugs. Identifying other antimalarials with the efficacy, tolerability and safety of the artemisinins will be a tall order, but the ever-growing threat of resistance to antimalarial drugs gives the work of Krishna and colleagues a practical significance beyond its undoubted academic merit.

Robert G. Ridley is at the UNDP/WorldBank/WHO Special Programme for Research and Training in Tropical Disease, World Health Organization, CH-1211 Geneva 27, Switzerland.
e-mail: ridleyr@who.int

- Haynes, R. *Curr. Opin. Infect. Dis.* **14**, 719–726 (2001).
- Eckstein-Ludwig, U. *et al.* *Nature* **424**, 957–961 (2003).
- Greenwood, B. & Mutabingwa, T. *Nature* **415**, 670–672 (2002).
- Ridley, R. G. *Nature* **415**, 686–693 (2002).
- Meshnick, S. R., Taylor, T. E. & Kamchonwongpaisan, S. *Microbiol. Rev.* **60**, 301–315 (1996).
- Robert, A., Cazes, J. & Meunier, B. *Chem. Int.* **40**, 1954–1957 (2001).
- Olliaro, P., Haynes, R. K., Meunier, B. & Yongyuth, Y. *Trends Parasitol.* **17**, 123–126 (2001).

Cell biology

Tumour jailbreak

Kenneth M. Yamada

New work shows that a cage-like matrix of protein fibres around cells can inhibit the growth of tumours. But cancer cells producing the enzyme MT1-MMP can cleave this matrix and proliferate freely.

Tumour cells resemble hardened criminals. They defy the body's social restraints, altering their behaviour and interactions to proliferate and spread as they please. The body also has potential physical restraints — the three-dimensional mesh-like matrices that provide support for normal cells — and Hotary *et al.*¹ now

report in *Cell* that these can prevent the proliferation of tumour cells in the laboratory. But if the cells are able to produce a specific matrix-cleaving enzyme, they can escape these physical restraints as well, becoming free to change shape and multiply.

One of the purposes of the 'extracellular matrix' is to provide structural support for

the cells that sit on top of it. When normal cells start on the path to malignancy, they can proliferate with ease on top of the matrix, in what is effectively a 'two-dimensional' environment. But once cancer cells start to invade the matrix in an attempt to spread to distant organs, they are hampered in several ways. First, because the matrix provides cells with growth and survival signals, which vary between tissues, tumour cells moving into a foreign tissue have to be able to adapt to the new signals². Second, the cells must be able to proliferate and expand the tumour mass within a resistant 'three-dimensional' matrix that is rich in fibres made of a protein called type I collagen. Third, the body often encases tumour cells in abnormal amounts of crosslinked fibrin protein, from which the cells need to escape³.

Given the defences provided by the extracellular matrix, it is perhaps not surprising that previous studies of cancer mechanisms have implicated a class of protein-cleaving enzymes, or proteases, known as the matrix metalloproteinases (MMPs). These enzymes are thought to be important for tumour-cell invasion and spread, by opening up paths in the matrix through which the cells can move. They have also been proposed to have indirect effects on cell proliferation, by freeing growth factors sequestered in the matrix and by promoting the growth of new blood vessels to supply the tumour with nutrients and oxygen⁴. It seemed unlikely that the MMPs would have a more direct effect on proliferation: studies of cells grown in regular two-dimensional culture systems (on top of matrix proteins in cell-culture dishes, for instance) have generally failed to find MMP-stimulated proliferation. But a rapidly growing body of literature has underscored the importance of the physical three-dimensionality of the matrix in regulating normal cell behaviour, including cell proliferation⁵⁻⁸. So Hotary and co-workers¹ compared tumour-cell proliferation in two-dimensional and three-dimensional conditions.

The authors found that the expansion of cell colonies inside three-dimensional collagen gels was suppressed. Yet the same cells could proliferate readily on the two-dimensional top of the gel. Vigorous growth inside the confines of such gels occurred only when the cells were engineered to produce one particular MMP — MT1-MMP — out of eight tested. Tumour cells producing large amounts of this enzyme could even grow well in the relatively dense connective tissue in the skin of mice. Yet MT1-MMP production conferred no growth advantages on cells on a two-dimensional substrate.

The enzyme, which is located in the cell membrane, needs its membrane 'anchor' and protease activity to permit growth under three-dimensional conditions. It seems to work by cleaving the surrounding collagen matrix (Fig. 1): when Hotary *et al.* grew

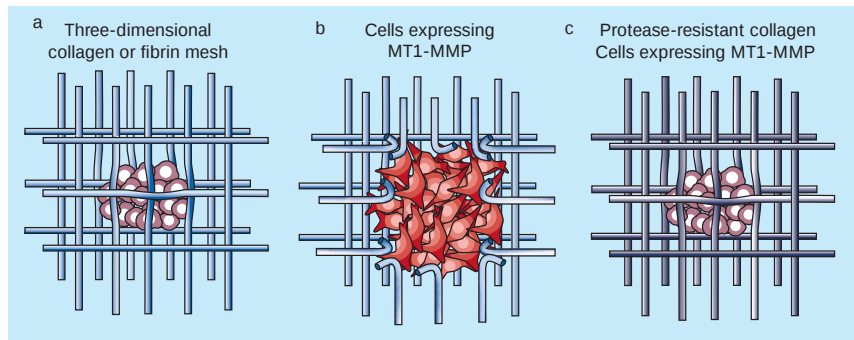


Figure 1 Tumour cells: prisoners and escapees. a, Hotary *et al.*¹ have found that cancer cells *in vitro* that are surrounded by a three-dimensional matrix of type I collagen or crosslinked fibrin are prevented from changing shape and proliferating. b, Production of the enzyme MT1-MMP by the cells results in local cleavage of matrix proteins; the cells then become free to change shape and so to proliferate. c, Using a collagen that cannot be cleaved suppresses proliferation, even in the presence of MT1-MMP. Likewise, treating MT1-MMP-producing tumour cells with the inhibitor TIMP-2 results in growth suppression in three dimensions (but not on flat surfaces — that is, in two dimensions).

MT1-MMP-producing cells in gels composed of a mutant collagen that could not be degraded, proliferation was blocked. Moreover, surrounding the tumour cells with crosslinked fibrin also suppressed growth unless MT1-MMP was active. These and other studies suggest that this protease is particularly effective in cleaving local collagen or fibrin, without help from other MMPs.

How does a three-dimensional matrix restrain tumour-cell proliferation? Previous work⁹ showed that cell shape — which is linked to the organization of the cell's internal 'skeleton' — is important in regulating proliferation. Hotary *et al.* found that tumour cells trapped in collagen or fibrin gels had an abnormally round shape, with actin filaments (a component of the cytoskeleton) being assembled only in the cell periphery. But cells producing MT1-MMP were able to stretch out and organize robust actin bundles. So the authors suggest that MT1-MMP promotes tumour-cell proliferation by disrupting an otherwise constricting three-dimensional matrix and permitting the changes in cell geometry that are needed for proliferation.

One of the most pressing questions raised by these exciting findings is whether membrane-type MMPs such as MT1-MMP are essential for the growth of human cancers. Further measurements of these enzymes in various tumours will reveal how often their levels are increased. Studies of the behaviour of tumours developed from cells genetically engineered to lack MT1-MMP or other membrane-anchored MMPs should likewise be enlightening. Also, some normal cells, such as human fibroblasts, can proliferate readily in certain three-dimensional matrices^{6,8}. It will be important to compare the effects of known regulatory properties of these matrices, such as internal tension and pliability, on normal and tumour cells.

Could these findings have implications for treating cancer? Initial clinical tests of

broad-spectrum MMP inhibitors have been disappointing⁴, even though they would have been expected to inhibit membrane-anchored MMPs (unless effective concentrations were toxic, causing side effects before they could be useful). Moreover, Wolf *et al.*¹⁰ have shown that cells in which all proteases are inhibited can still migrate in collagen (although the authors used a form of collagen that does not prohibit changes in tumour-cell shape). But Hotary *et al.*¹ point out that the chemical inhibitors used in such studies are not specific to MMPs, but also inhibit another major class of protease, the adamalysins. This additional inhibition might have inadvertently interfered with therapy. So it will be important to develop inhibitors of greater specificity, to establish the importance of membrane-type MMPs in cancer. If these enzymes are indeed crucial, then such inhibitors could also be useful therapeutically. Clinical experience shows how difficult it is to selectively kill every last cancer cell, by chemotherapy or other means. But a strategy of preventing the escape of any surviving cells from a matrix prison might just work.

Kenneth M. Yamada is at the National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, Maryland 20892-4370, USA.
e-mail: kenneth.yamada@nih.gov

- Hotary, K. B. *et al.* *Cell* **114**, 33–45 (2003).
- Frisch, S. M. & Screaton, R. A. *Curr. Opin. Cell Biol.* **13**, 555–562 (2001).
- Dvorak, H. F., Senger, D. R. & Dvorak, A. M. *Cancer Metastasis Rev.* **2**, 41–73 (1983).
- Coussens, L. M., Fingleton, B. & Matrisian, L. M. *Science* **295**, 2387–2392 (2002).
- Bissell, M. J. & Radisky, D. *Nature Rev. Cancer* **1**, 46–54 (2001).
- Cukierman, E., Pankov, R., Stevens, D. R. & Yamada, K. M. *Science* **294**, 1708–1712 (2001).
- Walpita, D. & Hay, E. *Nature Rev. Mol. Cell Biol.* **3**, 137–141 (2002).
- Grinnell, F. *Trends Cell Biol.* **13**, 264–269 (2003).
- Chen, C. S., Mrksich, M., Huang, S., Whitesides, G. M. & Ingber, D. E. *Science* **276**, 1425–1428 (1997).
- Wolf, K. *et al.* *J. Cell Biol.* **160**, 267–277 (2003).

Archaeology

Propaganda of the pyramids

Jared Diamond

You would have thought that massive monuments would be built by states at the apogee of their pomp and glory. Not so, according to an argument in which it is states on the up that need to impress.

It seems obvious that huge ancient monuments, such as the three Great Pyramids at Giza in Egypt, should be taken as testimony to a powerful ancient state behind them. As Joyce Marcus argues¹ in a contribution to a newly published book, that conclusion is indeed what the builders wanted us to draw — but it's the opposite of the truth. Often, the biggest monuments were built early in a state's history, before it reached its peak power. They were built for propaganda, precisely to conceal the state's lack of real clout.

For example, among those three pyramids at Giza, Khufu's was the largest ever erected in Egypt, with a base area of 5 hectares, a height of 146 metres, and a mass of 6 million tonnes. It may be the bulkiest single building erected in human history, and until recently it was the tallest. That is certainly not because Khufu commanded more resources than did any subsequent leader in history. Many governments today exercise far more power, but none would waste its resources on building a big pyramid just for display and lacking office space, auditoriums and theme restaurants. Khufu's dynasty came early in the Egyptian sequence; it was only the fourth of about 31 dynasties. Even Khufu's two successors didn't try to match him: Khafra's pyramid at Giza is slightly smaller, and Menkaura's is less than half of Khufu's in height. Later Egyptian dynasties drastically reduced the scale of their pyramids.

Yet those later dynasties were much more powerful than Khufu's. They chose to invest their power in other ways: launching long-distance trading expeditions and military campaigns of conquest, maintaining big garrisons, and building big fortresses, irrigation works and ship channels. All of those things were beyond Khufu's capabilities. As head of an early state, he couldn't dominate Egypt's neighbours. The Great Pyramid is a bluff, a proclamation of power as empty as that by Shelley's *Ozymandias*:

*'My name is Ozymandias, King of Kings:
Look on my works, ye Mighty and despair!
Nothing beside remains. Round the decay
Of that colossal wreck, boundless and bare,
The lone and level sands stretch far away.'*

Investment in architecture followed much the same course in the ancient Near World's two main centres of power, Mexico and Peru. Marcus points out that one of the earliest powerful states in the Valley of

Mexico arose at Teotihuacan, which built its largest structure, the Pyramid of the Sun, early in its history. Its base area of 5 ha matched that of Khufu's pyramid, although it was 'only' half as high. Teotihuacan continued to grow in population and in geographical extent of power, and the valley's later Aztec Empire was even more powerful, but neither again invested in such a large building. Instead, they poured resources into long-distance trade, outlying colonies, military conquests in the case of the Aztecs, garrisons, intensive agriculture and crafts production — much as had Khufu's successors in Egypt.

The story is similar for Peru's earliest state, that of the Moche, who in the centuries after AD 100 built Peru's largest pyramid, the Huaca del Sol, with a base area of the usual 5 ha but a height only one-fifth that of Khufu's pyramid. (Is there any reason why the largest pyramids of Egypt, Mexico and Peru all arrived at that same magic number for their base areas?) Peru's more powerful successor states of the Chimu and Incas, enjoying unquestioned actual control, evidently saw no need for ostentation. Instead, the Incas constructed a vast road system, storehouses and irrigation canals, sent armies far afield to conquer, and resettled subject peoples. Unable to do such things, the Moche made do with big but useless pyramids.

Tests of a good hypothesis are whether it can interpret cases other than those initially adduced as supporting evidence, and whether it can be broadened to encompass apparent exceptions. Since reading Marcus's article, I've kept asking myself about its relevance to sites that I visit as a tourist during overseas travels. I've come up with one more supportive example, and two cases of understandable exceptions.

When I visited Japan, I read that the emerging Japanese state had built huge keyhole-shaped earthwork mounds called kofun, and I asked my hosts to show me one. They took me to the largest of all kofun, that of the Emperor Nintoku near Osaka. At 32 ha, it covers much more ground than Khufu's pyramid, the Pyramid of the Sun or the Huaca del Sol, but it would have been easier to build because it consists just of earth rather than of stone, brick or adobe. Sure enough, I learned that Nintoku's and the other largest kofun were built early in the kofun era, when Japan's first state (the Yamato state) was just arising. It could only impress rather than conquer its neighbours,



100 YEARS AGO

Perhaps the phenomenon of mirage is not sufficiently rare in England to make its occurrence noteworthy, but I should like to mention a singularly beautiful example that I noticed on Sunday last... I was riding on my bicycle along the Upper Richmond Road towards the west, and against a fairly steady breeze, and had arrived at the part of the road lying between the railway bridge and the Putney High Street—about opposite house No. 110 — when I noticed that the road beyond, some fifty yards in front of me, was apparently flooded ankle deep in water. I was somewhat disconcerted at the prospect of riding through such a quantity of water, but I found to my astonishment that when I arrived at the supposed lake the road was perfectly dry. I thereupon turned and rode back to my previous station, and, dismounting, watched the phenomenon for some while. To assure myself that it was no personal illusion upon my part, I directed the attention of a passing stranger to the scene, and he was as impressed as I had been. I should mention that the road sloped slightly downhill from me, and the sun was high (12.50 p.m.) above on my left. From *Nature* 20 August 1903.

50 YEARS AGO

It has long been recognized that the forced ascent of air over mountains can produce cloud. Only in the past twenty years has it been realized that thin clouds 10,000 ft. or more above the mountains and with no cloud beneath them may owe their existence to the presence of the mountains. Such clouds are not formed by the direct lifting of the lower air up to cloud-level but in the ascending currents of a system of waves which can be produced by the mountain in somewhat the same way that a rock on a river bed produces waves downstream. The wind and temperature structures of the air have to fulfil certain conditions, as Dr. R. S. Scorer... has shown. The favourable conditions are an increase with height of both wind-speed and of the temperature lapse-rate. An inversion of temperature in the lower layers of the air followed by a rapid fall of temperature higher up is a favourable arrangement; but a steep fall of temperature with height low down, such as occurs on sunny afternoons, is unfavourable.

From *Nature* 22 August 1953.

and hadn't yet succeeded in expanding beyond its homeland in the Kinai region.

I then visited Easter Island, famous for its giant stone statues. There I learned that later statues are bigger than early ones, and that the tallest ever erected (the one named Paro, 9.8 m tall) was the last — in apparent disagreement with Marcus. But Easter Island, unlike Egypt or the Valley of Mexico or Peru, never became tightly unified and remained divided into rival clans that continued to compete visibly with each other. Hence Easter Island might violate the letter but supports the spirit of Marcus's hypothesis.

Finally, the Maya city-states in Central America are famous for their own pyramids and temples. As on Easter Island, later Maya rulers built bigger temples, but again the Maya states were never unified but stayed locked in fierce competition and warfare.

In addition, Marcus herself points out that some big late Maya buildings, such as Pacal's tomb at Palenque and Hasaw Chan K'awil's tomb at Tikal, were erected by usurpers or else by kings weaker than their predecessors, and thus with a special need to indulge in flashy displays of power.

Archaeologists studying other ancient monuments will find it challenging to test or expand Marcus's arguments. As she concludes, "We should be as skeptical of ancient propaganda as we are when dealing with modern politicians".

Jared Diamond is in the Departments of Geography and Environmental Health Sciences, University of California, Los Angeles, California 90095-1524, USA. e-mail: jdiamond@geog.ucla.edu

1. Marcus, J. in *Theory and Practice in Mediterranean Archaeology: Old World and New World Perspectives* (eds Papadopoulos, J. K. & Leventhal, R. M.) 115–134 (Cotsen Inst., Los Angeles, 2003).

Earth science

Tiny triggers deep down

Harry W. Green II

The documentation and characterization of remotely triggered earthquakes deep within the Earth is an achievement that provides insight into the mechanisms that initiate such events.

Earthquakes occur widely in the planet's crust and to depths approaching 700 km in subduction zones, where oceanic crust and the associated 50–100 km of mantle dive back into Earth as the return

flow of plate tectonics. But we know little about the physics of earthquake initiation (nucleation), especially at great depth, because the mechanisms known to operate close to the surface — brittle failure of

virgin rock, or frictional sliding on a pre-existing fault — cannot occur at the high pressures at depth¹.

A new window on the problem may have been opened by Tibi *et al.*² (page 921 of this issue). They provide the first analysis of two large (magnitudes 7.6 and 7.7), very deep earthquakes that occurred on 19 August 2002 in the Tonga subduction zone beneath the southwestern Pacific Ocean. Although these earthquakes were separated by about 300 km on the map and by 65 km in depth, they occurred within 7 minutes of each other. Tibi *et al.* argue that the second large earthquake, and a magnitude-5.9 precursor of it, were triggered by passage of the seismic waves generated by the first earthquake.

Although remote triggering is known for earthquakes near Earth's surface³, Tibi *et al.* provide the first such demonstration for a deep earthquake. Moreover, the authors discuss an earthquake series that occurred beneath Tonga in 1986 (Fig. 1) that now can be seen as probably another remotely triggered sequence. It is clear that regions in which earthquakes are triggered by the small disturbances generated by earthquake waves far from the source must be primed for failure, but for some reason nucleation does not occur readily. It is also clear that the delay between arrival of the triggering seismic waves and the time of the ensuing deep earthquakes varies from minutes to tens of minutes (see Table 1, page 922). Thus, the timescale of this 'incubation' period is likely to be a characteristic of the triggering mechanism.

Three mechanisms have been proposed as potentially responsible for deep earthquakes: (1) dehydration embrittlement⁴; (2) faulting induced by a phase transformation between one mineral form (olivine) and another, denser form (spinel)⁵; and (3) adiabatic shear instability⁶. All three have an experimental basis (although for crystalline materials, the last has been demonstrated only in metals). In each case, shear failure is the end result — rapid slip across a narrow zone such as a fault.

Mechanism 1 basically extends brittle fracture to high pressures by the generation of a pore fluid that assists opening of tensile microcracks, which then self-organize and lead to shear failure. Mechanism 2 is similar in outcome, but the underlying physics is fundamentally different. It involves the generation of another type of defect — microanticracks — which are small, crack-shaped lenses filled with a low-viscosity nanocrystalline aggregate of the stable phase; the microanticracks then self-organize and lead to shear failure⁷. Mechanism 3 involves the localization of deformation into a shear zone as a result of strain-softening: the rock becomes weaker as it flows. Runaway shear heating follows, leading to failure.

All three processes have specific requirements for generating an earthquake. The first requires a hydrous phase at or slightly beyond

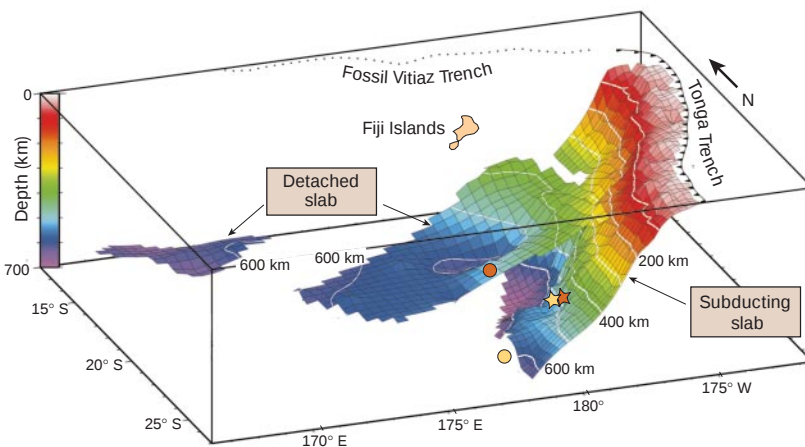


Figure 1 Earthquakes and subducted slabs beneath the Tonga–Fiji area. The subducting slab and detached slab are defined by the historic earthquakes in this region: the steeply dipping surface descending from the Tonga Trench marks the currently active subduction zone, and the surface lying mostly between 500 and 680 km, but rising to 300 km in the east, is a relict from an old subduction zone that descended from the fossil Vitiaz Trench. The locations of the mainshocks of the two Tongan earthquake sequences discussed by Tibi *et al.*² are marked in yellow (2002 sequence) and orange (1986 series). Triggering mainshocks are denoted by stars; triggered mainshocks by circles. The 2002 sequence lies wholly in the currently subducting slab (and slightly extends the earthquake distribution in it), whereas the 1986 mainshock is in that slab but the triggered series is located in the detached slab, which apparently contains significant amounts of metastable olivine^{8,9}. (Modified from ref. 13 with permission of the American Geophysical Union.)

its limit of stability that can break down to produce the fluid necessary to produce the instability. For the second, because of low temperatures in the core of the subducting slab, olivine must have failed to react to the spinel phase that is stable at depths of 400–700 km, and must be slowly transforming and causing earthquakes as it warms up¹. The third requires specific conditions for slow, continuing flow to be concentrated into a narrow zone in which the strain rate can accelerate as the heat generated by the straining accumulates, leading to an explosive increase in temperature, melting and shear failure. Each of these mechanisms has different implications for the temperature of subducting slabs and for the possible recycling of water back into the deep mantle from the surface.

Can the incubation times of the Tongan sequences help to discriminate between these possibilities? I think that they can. One cannot be certain whether the required phases are present for mechanisms 1 or 2, but we know from experimental work in the laboratory that a few minutes to tens of minutes is sufficient time in which to generate the primary microcracks or microanticracks, and for them to self-organize and lead to failure. In contrast, an adiabatic shear instability, in which strain rates are initially low, must inherently have a slow lead-up period as strain-induced heat accumulates to drive the rapid stage of the process. This time has been estimated as 10–10,000 years⁹. So unless the regions in which the earthquakes beneath Tonga were triggered contained shear zones that were close to thermal runaway, it is difficult to see how mechanism 3 could be responsible. This is particularly true for the events of August 2002, because the triggered earthquakes lie in a region in which an earthquake had never been detected previously (Fig. 1).

The 1986 Tongan sequence might tell us even more about deep-earthquake nucleation. As Tibi *et al.* point out², the triggering mainshock lay in the steeply dipping, currently active subduction zone, but the triggered earthquakes were in a remnant slab lying above it (Fig. 1). Various data^{9,10} are consistent with the presence of a significant amount of metastable olivine in this slab but are less consistent with other possibilities. Thermal models of subduction zones show that the currently active Tonga slab is the coldest on Earth and therefore has the highest probability that metastable olivine is preserved within it¹¹. Thus, if there is metastable olivine in the remnant slab, its presence in the active slab is virtually assured, which could be responsible for the initiation of all of the earthquakes in these sequences — although, after initiation, it is possible that adiabatic shear heating could contribute to the total size and magnitude of the earthquakes¹². In contrast, hydrous phases are difficult to reconcile with the properties of the detached slab beneath Fiji^{8,9,13}, and more generally it is not clear that they

can trigger earthquakes at depths of more than 400 km (ref. 1).

Tibi and colleagues' observations are a major advance in understanding deep earthquakes, and they might provide a new constraint on the mechanism by which these earthquakes begin. This long-standing problem in geophysics is far from solved, however. Further searches for other triggered sequences of deep earthquakes, and for the possible existence of metastable olivine and/or hydrous phases, will be necessary for us to take the next steps in understanding. ■

Harry W. Green II is at the Institute of Geophysics and Planetary Physics, and the Department of Earth Sciences, University of California, Riverside, California 92521, USA.
e-mail: hgreen@mail.ucr.edu

1. Green, H. W. II & Marone, C. J. in *Plastic Deformation of Rocks* (eds Wenk, H. R. & Karato, S.) 181–199 (Mineralogical Society of America, Washington DC, 2002).
2. Tibi, R., Wiens, D. A. & Inoue, H. *Nature* **424**, 921–925 (2003).
3. Gombert, J. J. *Geophys. Res.* **101**, 751–764 (1996).
4. Raleigh, C. B. & Paterson, M. S. *J. Geophys. Res.* **70**, 3965–3985 (1965).
5. Green, H. W. II & Burnley, P. C. *Nature* **341**, 733–737 (1989).
6. Griggs, D. T. & Baker, D. W. in *Properties of Matter Under Unusual Conditions* (eds Mark, H. & Fernbach, S.) 23–42 (Wiley, New York, 1969).
7. Green, H. W. II *Sci. Am.* **271**, 64–71 (1994).
8. Ogawa, M. *J. Geophys. Res.* **92**, 13801–13810 (1987).
9. Chen, W.-P. & Brudzinski, M. R. *Science* **292**, 2475–2479 (2001).
10. Chen, W.-P. & Brudzinski, M. R. *Geophys. Res. Lett.* doi:10.1029/2002GL016330 (2003).
11. Mosenfelder, J. L. *et al. Phys. Earth Planet. Inter.* **127**, 165–180 (2001).
12. Kanamori, H., Anderson, D. L. & Heaton, T. H. *Science* **279**, 839–842 (1998).
13. Brudzinski, M. & Chen, W.-P. *J. Geophys. Res.* doi:10.1029/2002JB002012 (2003).

Developmental biology

Hotspots for evolution

Michael K. Richardson and Paul M. Brakefield

Two studies of fruitflies suggest that although development relies on a diverse toolkit of genes, the evolution of physical characteristics might be powered by variation in just a few of these tools.

The physical characteristics of animals evolve because their genes change over successive generations. It is not always clear, though, which genes are involved^{1,2}. The genes that regulate embryonic or larval development are likely candidates, because they control how the animal develops its characteristic form and features. It is possible that natural selection might produce evolutionary change after the adjustment of just a few such switches on the genetic control panel of development. Writing in this issue, Sucena *et al.*³ and Gompel and Carroll⁴ provide evidence that this can indeed happen. They show that modifications at just a few developmental hotspots underlie 'parallel' evolutionary changes that occurred independently in different species.

Tinkering with developmental genes is not necessarily an easy route to evolutionary change. For example, mutations in the *antennapedia* gene — an important regulator of development in the fruitfly *Drosophila* — can produce a fly with legs growing on its head⁵. This may be fascinating to developmental geneticists, but from the fly's point of view it is not helpful.

So, how are developmental genes altered during the normal course of evolution in natural populations? To answer this question, researchers need to marry a knowledge of evolutionary changes with developmental genetics — as Sucena *et al.*³ and Gompel and Carroll⁴ have now done. Both groups had already made preliminary studies with a single species of the fruitfly *Drosophila melanogaster* (one of the key model species in developmental

genetics). But model species alone cannot tell us everything about evolution, and so each team broadened their investigations to include other fruitflies with distinctive physical characteristics (morphology). They then searched for changes in genetic pathways that might account for the differences.

Sucena and colleagues³ examined the development of hairs — called trichomes — in the larvae of different *Drosophila* species. They found that hairless patches on the young larvae had evolved independently in three of the lineages included in the study (Fig. 1). If something evolves once, it can be difficult to find out why, but if it evolves three times independently within a species group, we can look for correlations by mapping developmental changes and trait evolution onto a 'phylogenetic' tree (a sort of family tree)⁶. And Sucena and colleagues found such a correlation: the activity of a gene called *shavenbaby* (*svb*) was absent from the naked areas of all three lineages that showed hair loss.

These findings could mean that a loss of *svb* expression was directly responsible for the trichome loss. Alternatively, the loss of *svb* could simply be a consequence of another, more important, change that occurred earlier in development. But, as Sucena and colleagues note, previous studies have shown that several 'upstream' genes involved in trichome patterning, including *wingless* and *engrailed*, are not altered in certain species, related to *Drosophila virilis*, that show trichome loss^{7,8}. To determine exactly how the *svb* gene affected trichome development, the authors used another powerful tool — they crossed certain species

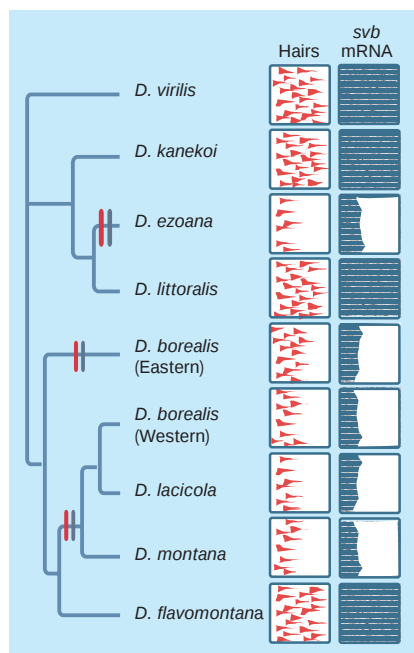


Figure 1 Shavenbabies are hairy. As described by Sucena *et al.*³, the correlation between the loss of trichomes (hairs) in different fruitfly species and the loss of the activity of the *shavenbaby* gene (*svb*) is revealed when the data are mapped onto a family tree (a phylogeny). The boxes represent trichome coverage and *svb* expression (assessed through measuring levels of messenger RNA, mRNA) in an anterior segment of the fruitfly's abdomen. In all three lineages in which trichome loss arose independently, *svb* expression is also lost (vertical red and grey bars, respectively, on the tree). Gompel and Carroll⁴ offer a similar but more detailed picture of their results in their Fig. 2 (page 933).

of *Drosophila* and looked at gene activity in hybrid animals. These additional data enabled them to infer that a loss of *svb* expression itself was probably directly responsible for the evolution of hair loss.

Meanwhile, Gompel and Carroll⁴ surveyed variation across species in the abdominal pigmentation pattern of adult fruitflies. Each segment of the abdomen in *D. melanogaster* bears a dark band. The sexes also differ: the two terminal segments are darkly pigmented in males, but not in females. This patterning is controlled by the products of two *bric-a-brac* genes, *bab1* and *bab2*. The Bab proteins repress pigmentation and are produced in segment-specific and sex-specific patterns in the pupal abdomen during development. In *D. melanogaster*, Bab function is required not only for pigmentation but also for trichome development.

The authors analysed 13 species of *Drosophila* and found several different patterns of *bab2* expression. In most of the species, patterns of *bab2* activity were well correlated with the pigmentation of the adult flies, suggesting that, in these animals, Bab2 does

indeed function as a regulator of pigmentation and is involved in the evolution of similar pigment changes in different lineages. But in some species, *bab2* activity was correlated not with pigmentation but instead with the pattern of trichomes on the abdomen. So although pigmentation and trichome patterning changes are often coupled, their evolution has become uncoupled in some groups. Gompel and Carroll suggest that the two traits become uncoupled when they are under different selective pressures. For example, when pigmentation changes, but trichomes do not, the biological function of the latter might somehow be important for survival.

Both studies suggest that although many genes are involved in the development of physical characteristics, some evolutionary changes — including examples of convergence, in which independent evolutionary events in different species result in similar physical characteristics — involve key regulatory points. The *bab2* and *svb* genes might each function as convenient switches for modifying, or turning on and off, a given trait during evolution. Each gene might represent a developmental 'hotspot' for evolution — or, as Sucena *et al.* put it, a regulator that preferentially accumulates evolutionary change. Another possible example of a developmental hotspot can be envisaged during vertebrate limb evolution: changes to genes that determine the duration of limb growth in the embryo might have been important in the evolution of specialized limb types, such as the dolphin flipper⁶.

Although not all instances of trait convergence rely on the same genetic mechanism, these studies have uncovered several instances that do. And this creates a headache for biologists struggling with the concept of homology (in developmental genetics, homologous characteristics are defined as being 'identical by descent', which means they are derived from equivalent genetic networks in common ancestors rather than arising independently in separate lineages). Loss of trichomes in different lineages is traditionally considered an example of parallel evolution and not true

homology, because the loss was not inherited — instead, it arose independently in the different species. But if changes in the same gene are involved in each lineage, we could perhaps say that trichome loss is homologous at the level of the underlying process.

The activity of *bab2* and *svb* has been modified repeatedly in evolution, probably largely through changes in the regulatory sequences that switch the genes on or off. Understanding how such modifications have occurred might reveal why some genes and not others are developmental hotspots for driving evolutionary change. Sucena *et al.* suggest that *svb* might be a hotspot because it integrates numerous genetic inputs to control the final output — the pattern of hairs⁵.

The two new studies^{3,4} have revealed some of the genetic mechanisms that might underlie the evolution of hair loss and pigmentation. But they also point to a future challenge — understanding the function of the physical characteristics, and the causes of the evolutionary change. The integration of evolution with development is beginning to encompass the broad tools of quantitative genetics and functional genomics. This approach will show us how physical changes are generated, but it will tell us nothing about why evolutionary change has taken place in natural environments⁹. The brightest future for evolutionary developmental biology might lie instead with the study of systems in which we can analyse the cause as well as the function of evolutionary change. ■

Michael K. Richardson and Paul M. Brakefield are at the Institute of Biology, Leiden University, PO Box 9516, 2300 RA Leiden, The Netherlands. e-mails: richardson@rulsfb.leidenuniv.nl
brakefield@rulsfb.leidenuniv.nl

1. Stern, D. L. *Evolution* **54**, 1079–1091 (2000).
2. Tautz, D. & Schmid, J. *Phil. Trans. R. Soc. Lond. B* **353**, 231–240 (1998).
3. Sucena, E., Delon, I., Jones, I., Payre, F. & Stern, D. L. *Nature* **424**, 935–938 (2003).
4. Gompel, N. & Carroll, S. B. *Nature* **424**, 931–935 (2003).
5. Struhl, G. *A. Nature* **292**, 635–638 (1981).
6. Richardson, M. K. & Chipman, A. D. *J. Exp. Zool.* **296B**, 8–22 (2003).
7. Bokor, P. & DiNardo, S. *Development* **122**, 1083–1092 (1996).
8. Hatini, V. *et al. Genes Dev.* **14**, 1364–1376 (2000).
9. Beldade, P. & Brakefield, P. M. *Nature Rev. Genet.* **3**, 442–452 (2002).

Mathematics

The 24-dimensional greengrocer

Ian Stewart

The best way to stack oranges has been evident in markets around the world for centuries, but the mathematics of the problem is far from trivial. The solution for the 24-dimensional case is now within reach.

Put a handful of identical coins on the table, and push them around until they fit together as closely as possible. You will get a honeycomb pattern, or hexagonal lattice, in which each coin is tightly surrounded by six others (Fig. 1, overleaf). This experiment suggests two things: that

the 'kissing number' in two dimensions is six, and that the hexagonal lattice is the most efficient way to pack circles. The Greeks could have proved the first statement with complete logical rigour, had they thought to do so. The second statement, though widely suspected, was not properly



Figure 1 Kissing... coins. In this two-dimensional example of packing, each coin is surrounded by six others. In fact, this is the closest packing arrangement possible, and defines the 'kissing number' for two dimensions. Cohn and Elkies¹ have now tied down the kissing number for spaces with 8 and 24 dimensions.

proved until 1940, which demonstrates how tricky this area of mathematics can be.

The problem can be generalized to the packing of spheres (strictly, many-dimensional hyperspheres) in higher-dimensional spaces. This seemingly esoteric field of research has important applications in electronic communications, because it is fundamentally linked to digital codes. Currently, we know the kissing number for spaces of dimension 1, 2, 3, 8 and 24 — and no others. We are also pretty sure what the most efficient packing looks like in those dimensions, but we can prove that we're right only in dimensions 1, 2 and 3. Thanks to advances in technique introduced by Henry Cohn and Noam Elkies, reported in *Annals of Mathematics*¹, we are now very close to polishing off dimensions 8 and 24 as well. Better still, we know a lot more about all the other dimensions.

It all began in 1611, when Johannes Kepler wrote a book as a New Year's gift to his sponsor: *De Nive Sexangula* (*The Six-Cornered Snowflake*). Wondering why many snowflakes have six-fold symmetry, Kepler thought his way to a version of the atomic lattice of a crystal: a tightly packed assembly of identical units. He described three ways to pack spheres in space, and remarked offhandedly that for one of them, the way greengrocers pile oranges, "the packing will be the tightest possible" (Fig. 2). Only in 1998 did Thomas Hales² prove that Kepler was right.

During the intervening 387 years, mathematicians proved ever-better estimates of the packing density. Specific packings lead to lower bounds (the packing can be at least this tight). Upper bounds (the density cannot be as big as this) are trickier. When the upper and lower bounds get close together, the problem is under control; when they become equal, it is solved. Until this year, the best-known upper bound for packings in 8 dimensions was 1.012 times the lower bound; in 24 dimensions the ratio was 1.27. In most other dimensions the discrepancy was far worse. These ratios have now been improved to 1.000001 and 1.0007, respectively.



Figure 2 Orange arrangement. Greengrocers do it instinctively, but mathematically it is proven that the closest packing in three dimensions is achieved with a kissing number of 12: each orange enclosed inside this stack is in contact with 12 neighbours.

Why are 8 and 24 so special? It all seems to revolve around a bizarre coincidence. There is a remarkable formula in number theory that can sometimes provide good upper bounds for lattice packings; it works best in dimensions 8 and 24. In these dimensions there are also some very special lattices, called the E_8 root lattice and the Leech lattice, which provide unusually good lower bounds. As it happens, the upper and lower bounds are very close together — close enough that we can say, with complete confidence, that in 24 dimensions the kissing number is exactly 196,560. That is, you can pack that many unit

spheres, and no more, so that they all touch a given sphere. In 8 dimensions the kissing number is exactly 240. In 1, 2 and 3 dimensions it is 2, 6 and 12. But, for example, in 7 dimensions the best we can say is that the kissing number lies between 126 and 140.

The new work starts from the connection between sphere-packings and binary codes. A string of n binary digits can be considered as a point in a space of n dimensions; strings are very similar if these points are close together. Usually, new methods of packing spheres have been used to determine new codes, but Cohn and Elkies¹ play the game in the opposite direction, using ideas about codes to estimate packing densities for spheres. Their main technique is a version of linear programming, an optimization method widely used in the commercial world. In all dimensions, they obtain improved bounds. In dimensions 8 and 24, the improvement is so striking that they conjecture that similar techniques will prove that the E_8 and Leech lattices provide the tightest possible packings in those dimensions. Kepler would have been delighted. ■

Ian Stewart is at the Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK.
e-mail: ins@maths.warwick.ac.uk

1. Cohn, H. & Elkies, N. *Ann. Math.* **157**, 689–714 (2003).

2. Hales, T. Preprint at <<http://arxiv.org/abs/math.MG/9811078>> (1998).

Cell biology

Metabolism meets death

Julian Downward

A protein that controls cell death has been found in a complex with a protein involved in glucose metabolism. Is this a point of contact between these two crucial cellular processes?

All organisms, and all cells, need food to survive: it supplies energy for everything from movement to cell proliferation. But beyond this simple relationship, there is evidence that there is a more direct connection between the molecular pathways that convert food into energy within a cell and the pathways that control whether the cell lives or dies. For instance, the continued survival of mammalian cells depends on the availability of extracellular growth-factor proteins, which inhibit cell-death pathways¹ — and many of these proteins affect metabolism as well. Moreover, cancer cells not only have an increased metabolic rate but are also less likely to commit suicide when damaged². Some of the molecular links between cell death and metabolism are beginning to be fleshed out. On page 952 of this issue, Danial and colleagues³ propose another connection: they suggest that these processes are coordinately regulated through a killer protein called BAD.

Cell suicide is an everyday event in multicellular organisms and occurs by a process called apoptosis. One other hint of a relationship between apoptosis and metabolism was that, in mammalian cells, mitochondria — the compartments in which the products of the initial metabolism of glucose (glycolysis) are used to produce energy — have a key role in apoptosis. Many apoptotic stimuli cause the release of cytochrome *c* and other factors from mitochondria, which in turn trigger the activation of a group of proteins called caspases; these enzymes then dismantle the cell from within in an orderly manner⁴.

What determines whether this chain of events is activated? Proteins of the BCL-2 family are key regulators, with different members being either pro-apoptotic (such as BAD and BAX) or anti-apoptotic (such as BCL-2 itself), either promoting or preventing the release of cytochrome *c* respectively. BAD has been of particular interest. When cells are treated with growth or survival factors,

several enzymes are activated and in turn label BAD with phosphate groups. The result is that this newly phosphorylated protein is sequestered into inactive complexes that can no longer promote cell death⁵. So BAD has been proposed to be a point at which 'survival' signalling pathways intersect the machinery of death.

Might it also provide a point of connection between these events and cellular metabolism? Danial and colleagues' findings³ suggest that it might (Fig. 1). The authors took the approach of purifying the protein from mouse liver mitochondria to see which other proteins it might be interacting with. They found that BAD is present in large complexes containing several proteins that are likely to be involved in regulating its phosphorylation state, namely protein kinase A (PKA), protein phosphatase 1, and a PKA-anchoring protein, WAVE-1. But it was the fifth member of the complex that provided the real surprise: it was glucokinase, a member of the hexokinase family. These enzymes are responsible for phosphorylating glucose to produce glucose 6-phosphate — the first step in several pathways of glucose metabolism, including glycolysis and the storage of excess glucose as glycogen.

At this point more mature readers might find themselves reaching for dusty undergraduate biochemistry textbooks, wondering what on earth a killer like BAD is doing consorting with a regulator of glucose metabolism. In fact, the purchase of up-to-date copies of Stryer and of Lehninger by laboratories working on apoptosis has been gathering pace over the past few years, for several papers have reported close links between

apoptosis and glucose metabolism. In particular, the protein kinase Akt is activated by many growth and survival factors. It is also strongly anti-apoptotic, at least in part because it can phosphorylate and inactivate BAD⁶. And it has profound effects on cellular metabolism, stimulating glucose transport into the cell, the recruitment of hexokinases to mitochondria, and glycolysis^{7–9}. Most interestingly, Akt's ability to protect cells from apoptosis has been found to depend on the availability of glucose.

Why is there this requirement for glucose? This is not yet clear, although it has been reported that Akt promotes the interaction of hexokinase with the voltage-dependent anion channel (VDAC), which is needed to pump newly synthesized ATP (the end result of glucose metabolism, and the main form of readily available stored energy in the cell) out of mitochondria⁷. This interaction might prevent the channel from binding to BAX¹⁰. And because hexokinase is regulated by the level of glucose metabolites, its interaction with VDAC could perhaps respond to glucose levels. BAX is thought to be the ultimate regulator of cytochrome c release, and some data suggest that its direct binding to VDAC is essential for such regulation¹¹ — although this is controversial.

And what of BAD? Danial *et al.*³ report that when this protein is missing from cells, mitochondrial glucokinase activity and respiration (leading to ATP production) decrease. This hints that BAD is needed for the formation of the mitochondrially located complex that includes glucokinase, and that BAD thereby has a positive influence on metabolism.

Notably, its ability to form such a complex does not depend on its phosphorylation state. But the authors find that glucose promotes the phosphorylation of BAD, and that the glucokinase associated with phosphorylated BAD has higher activity than that associated with unphosphorylated BAD. So glucose-mediated phosphorylation of BAD might enhance glycolysis, as well as preventing cell death. That idea is supported by the authors' finding that mice that either lack BAD or express a non-phosphorylatable form of the protein show abnormal regulation of blood glucose. The defects are similar to those seen in mice with a liver-specific deletion of glucokinase, and also in people with maturity-onset type 2 diabetes of the young, a form of non-insulin-dependent diabetes mellitus in which glucokinase mutations have been found¹².

So the authors propose not only that BAD influences glucose metabolism but also that it is ideally positioned to serve as a 'sentinel', responding to abnormalities in glucose metabolism by triggering apoptosis. At present, however, it is not clear that the BAD–glucokinase complex reported here has any connection to the dependence of survival signalling on glucose that has been reported by others^{7–9}. Moreover, glucokinase is unique in the hexokinase family in that its expression is largely restricted to the liver and pancreatic β -cells. So it is unlikely to have a role in connecting apoptotic and metabolic pathways in all cell types. And glucokinase functions in the liver principally to drive glycogen synthesis, not glycolysis.

It will be interesting to see whether other, more broadly expressed hexokinases are found in complexes with BAD and, if so, what aspect of glucose metabolism they are involved in — or whether they simply act as glucose sensors. Danial and colleagues' report raises more questions than it answers, but it certainly provides new pointers as to where work in this area should be directed. The intimate connections between the control of cell death and metabolism are beginning to come to light, engendering renewed respect for old biochemical pathways.

Julian Downward is at the Cancer Research UK London Research Institute, 44 Lincoln's Inn Fields, London WC2A 3PX, UK.

e-mail: downward@cancer.org.uk

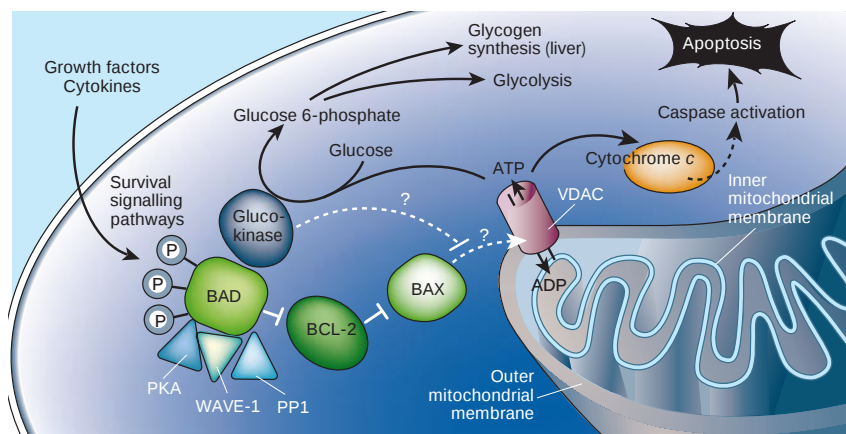


Figure 1 A point of integration for glucose metabolism and cell death? Danial *et al.*³ have found the pro-apoptotic protein BAD in a mitochondrially located complex with glucokinase (as well as protein kinase A, PKA; protein phosphatase 1, PP1; and WAVE-1). Quite what the purpose of the BAD–glucokinase link might be is unclear, but these proteins are known to influence apoptosis and metabolism, respectively. BAD antagonizes BCL-2, which in turn keeps the pro-apoptotic protein BAX in check. Survival signalling pathways phosphorylate BAD and prevent it from neutralizing BCL-2. BAX might interact with the voltage-dependent anion channel (VDAC), part of the channel that exchanges newly synthesized ATP for ADP⁷. This interaction might promote the release of cytochrome c from mitochondria¹¹, resulting in the activation of caspases, which in turn dismantle the cell. Relatives of glucokinase inhibit the interaction¹⁰. The metabolic role of glucokinase is to phosphorylate glucose to generate glucose 6-phosphate. In the liver, this is used to generate glycogen when blood sugar levels are high.

1. Raff, M. C. *Nature* **356**, 397–400 (1992).
2. Hanahan, D. & Weinberg, R. A. *Cell* **100**, 57–70 (2000).
3. Danial, N. N. *et al.* *Nature* **424**, 952–956 (2003).
4. Green, D. R. & Evan, G. I. *Cancer Cell* **1**, 19–30 (2002).
5. Puthalakath, H. & Strasser, A. *Cell Death Differ.* **9**, 505–512 (2002).
6. Datta, S. R., Brunet, A. & Greenberg, M. E. *Genes Dev.* **13**, 2905–2927 (1999).
7. Gottlob, K. *et al.* *Genes Dev.* **15**, 1406–1418 (2001).
8. Vander Heiden, M. G. *et al.* *Mol. Cell Biol.* **21**, 5899–5912 (2001).
9. Plas, D. R., Talapatra, S., Edinger, A. L., Rathmell, J. C. & Thompson, C. B. *J. Biol. Chem.* **276**, 12041–12048 (2001).
10. Pastorini, J. G., Shulga, N. & Hoek, J. B. *J. Biol. Chem.* **277**, 7610–7618 (2002).
11. Scorrano, L. & Korsmeyer, S. J. *Biochem. Biophys. Res. Commun.* **304**, 437–444 (2003).
12. Postic, C. *et al.* *J. Biol. Chem.* **274**, 305–315 (1999).

Developmental biology

Control stems from microRNAs

Dev. Cell **5**, 351–358 (2003)

According to Hristo B. Houbaviy *et al.*, the ability of embryonic stem (ES) cells to produce virtually any body tissue may be partly controlled by strings of 20–24 nucleotides, called microRNAs (miRNAs).

By screening mouse ES cells for miRNAs, the authors identified 12 miRNA-encoding genes that are unlike any previously found. Six of these genes, which lie in a cluster in the mouse genome, are switched off when the ES cells differentiate into balls of tissue called embryoid bodies.

Although the exact function of these six miRNA genes is unknown, Houbaviy *et al.* suggest that they may turn off other genes that drive ES-cell differentiation. The miRNAs are not found in adult mammalian tissues so far examined; the group has not yet looked for them in other stem cells.

The authors also show that humans have three miRNA genes that are related to the mouse cluster. These and other genes that help ES cells retain their remarkable developmental versatility are of interest to researchers trying to manipulate stem cells for clinical applications.

Helen Pearson

Archaeology

The dating game

New J. Phys. doi:10.1088/1367-2630/5/1/399 (2003)

The anti-corrosive properties of lead have made it a much-used material since antiquity — for pipes, for example. But the metal does degrade, if only very slowly, and S. Reich and colleagues have exploited that characteristic to propose a new dating method for archaeological artefacts.

Their method depends on the Meissner effect, the expulsion of magnetism when a metal becomes superconducting at very low temperatures. Lead becomes superconducting at 7.2 K. But the products of corrosion — principally lead oxide and lead carbonate — do not. Magnetic measurements can provide a figure for the mass of undecayed lead in an object, and if the total mass is known, the amount of corrosion product present can be estimated. Reich *et al.* tested this thinking on samples of objects of known age, from 2,500 years old (pictured) to 10 years old, and found that the mass of corrosion product is a direct reflection of how old they are.

The authors point out that the technique is non-destructive, as there is no need to separate the corroded content of an artefact from the pure metal. So, for instance, it could be used on coins made from another metal but with some lead content. The method depends on burial conditions remaining



Time piece — 2,500-year-old lead remains that have been sampled by Reich *et al.*

stable, however, especially pH. Lead degrades at slow and constant rates at pH values above 6.5, but much more quickly in more acidic environments.

Tim Lincoln

Cancer

Hostile takeover

Cell **114**, 323–334 (2003)

The cyclin D1 protein is frequently found at high levels in a wide range of human tumours. Justin Lamb and colleagues propose a mechanism by which this may turn a normal cell into a malignant one.

Cyclin D1 stimulates cell division. It has often been assumed that at high levels it causes cells to divide uncontrollably and thus promotes tumour formation. But analyses of tumour material failed to deliver consistent evidence to support this idea. Cyclin D1 also affects proteins that regulate gene transcription, although the data left room for doubt. But it is precisely this activity that now appears to turn a good cell bad.

Lamb *et al.* compared gene expression patterns of hundreds of human tumour samples, and uncovered a gene expression 'signature' associated with elevated levels of cyclin D1: the same 20 or so genes are turned on in many of the tumours. The authors further show that the protein C/EBP β , which has not previously been connected to cyclin D1, normally represses the expression of the signature genes. But cyclin D1 counteracts C/EBP β , and at sufficiently high levels overrides its repressive activity. Lamb and colleagues propose that the gene products thereby unleashed engage in a hostile takeover of the cell, resulting in cancerous growth.

Marie-Thérèse Heemels

Nonlinear physics

A brownian motor

Phys. Rev. Lett. **91**, 060602 (2003)

Thermal ratchets put noise to use by converting the randomness of thermal fluctuations to directional motion. This

process has been proposed as a mechanism by which motor proteins transport molecules within biological cells, and it has been demonstrated as a way of transporting small particles. Andreas Engel and colleagues give the topic a new twist by harnessing brownian rotation of nanoscale particles suspended in a fluid to produce a macroscopic torque on the fluid.

Although at face value it seems unlikely, a thermal ratchet breaks no thermodynamic laws. The directionality comes from some asymmetry imposed on the system, for example particles displaying brownian motion in an asymmetric field. In the present case, the nanoparticles are ferromagnetic: the suspension is a ferrofluid. And the asymmetry is supplied by an anharmonically oscillating magnetic field, acting in conjunction with a static field. The oscillating field imposes a rotational bias on the particles, which makes the carrier fluid flow by hydrodynamic coupling. So a 16-mm ferrofluid-filled sphere hanging by a thread in the magnetic fields begins to rotate — a kind of noise-powered motor.

Philip Ball

Neurobiology

Untangling tau

J. Neurosci. **23**, 6972–6981 (2003)

Hibernating animals that are in torpor — an inactive, hypothermic state — endure a rapid, reversible decline in the number of connections between certain types of nerve cell. Thomas Arendt and colleagues now show that this impressive structural plasticity is accompanied by changes in the phosphorylation of tau, a protein more commonly associated with neurodegenerative disorders.

Highly phosphorylated tau is the main component of the 'neurofibrillary tangles' that appear in the brains of patients with Alzheimer's disease and related conditions. The build-up of tangles is particularly pronounced in brain regions that enjoy a relatively high degree of structural plasticity, such as the hippocampus. But the process of tau phosphorylation and its potential involvement in neurodegeneration are not well understood.

Arendt *et al.* looked at the phosphorylation of tau in the hippocampi of ground squirrels during hibernation. In a region known as CA3, they found that synapses — the contacts between nerve cells — seemed to regress during torpor, but were re-established during periods of arousal. The decrease in connectivity was paralleled by the formation of highly phosphorylated tau.

The authors hope that hibernation will be a valuable model for studying the regulation of tau phosphorylation, and its potential link to synaptic plasticity and neurodegeneration.

Rebecca Craven

Fibre-optical features of a glass sponge

Some superior technological secrets have come to light from a deep-sea organism.

Modern technology cannot yet compete with some of the sophisticated optical systems possessed by biological organisms^{1–3}. Here we show that the spicules of the deep-sea 'glass' sponge *Euplectella* have remarkable fibre-optical properties, which are surprisingly similar to those of commercial telecommunication fibres — except that the spicules themselves are formed under normal ambient conditions and have some technological advantages over man-made versions.

The skeleton of the hexactinellid class of sponges is constructed from amorphous, hydrated silica^{3–6}. *Euplectella* builds an intricate cage (Fig. 1a), which typically houses a mating pair of shrimp (hence its nickname, 'Venus flower-basket') and is composed of a lattice of fused spicules⁴ that provide extended structural support.

A network of anchorage spicules (basalia) extend outwards in a crown-like formation. These spicules are generally 5–15 cm long and 40–70 μm in diameter; their native cross-section is homogeneous and they have no structural boundaries. Under stress or etching, the spicules reveal a characteristic layered morphology⁶ and cross-sectional variations in composition that appear as three distinct regions: a pure silica core of about 2 μm in diameter that encloses an organic filament; a central cylinder that has the greatest organic content of the three; and a striated shell that has a gradually decreasing organic content and which is glued together by organic films (Fig. 1b).

We anticipated that the spicules' rich substructure should be reflected in their optical properties as well. Indeed, interferometric refractive-index profiling⁷ revealed three regions that correspond to the three regions of structural composition (Fig. 1c): a core with high refractive index that is comparable to (or higher than) that of vitreous silica; a cylinder of lower refractive index that surrounds the core; and an oscillating pattern with progressively increasing refractive index at the outer part of the spicule.

To determine whether this typical 'core-cladding' refractive-index profile endows the spicules with wave-guiding properties, we investigated their transmission characteristics. We found that embedded spicules act as single- or few-mode waveguides — that is, light waves are effectively confined to the core, where refractive index is highest (Fig. 1d, left). When light was coupled into free-standing spicules, they functioned as multi-mode fibres, with most of the light filling the entire cladding, because of the enhanced refractive-index contrast

between the spicule and air (Fig. 1d, right).

These biological fibres therefore resemble commercial telecommunication fibres, in that they are made of the same material and have comparable dimensions, as well as similar refractive indices for the high-index core and a low-index cladding. They also function as efficient single-mode, few-mode or multi-mode waveguides, depending on the optical launch conditions.

The principal weakness of commercial optical fibres is that they fracture as a result of crack growth, whereas the spicules' lamellar layers, connected by organic ligands at the fibre's exterior, provide an effective crack-arresting mechanism and enhance fracture toughness^{6,8,9}. Another superior feature of the spicules is their formation under ambient conditions, a process that is regulated by organic molecules^{10,11}. This ambient-temperature process, unlike the high-temperature manufacture of man-made fibres, allows the structure to be doped with specialized impurities that improve the refractive index and therefore the wave-guiding properties. Our preliminary elemental analysis

shows, for example, that sodium ions are present throughout the spicules, particularly in the core. Although sodium ions (and many other additives) are desirable fibre-optic dopants, they present a manufacturing challenge, for example by causing devitrification at high temperatures.

Our results suggest the intriguing possibility that the spicules of *Euplectella*, beyond structural anchorage support, could also provide a highly effective fibre-optical network, which may be useful in distributing light in its deep-sea environment. This illuminating sponge should also shed light on low-temperature, biologically inspired processes that could give rise to better fibre-optical materials and networks.

Vikram C. Sundar*, **Andrew D. Yablont†**, **John L. Grazul***, **Micha Ilan‡**, **Joanna Aizenberg***

*Bell Laboratories/Lucent Technologies, Murray Hill, New Jersey 07974, USA
e-mail: jaizenberg@lucent.com

†OFS, Murray Hill, New Jersey 07974, USA

‡Department of Zoology, Tel Aviv University, Tel Aviv 69978, Israel

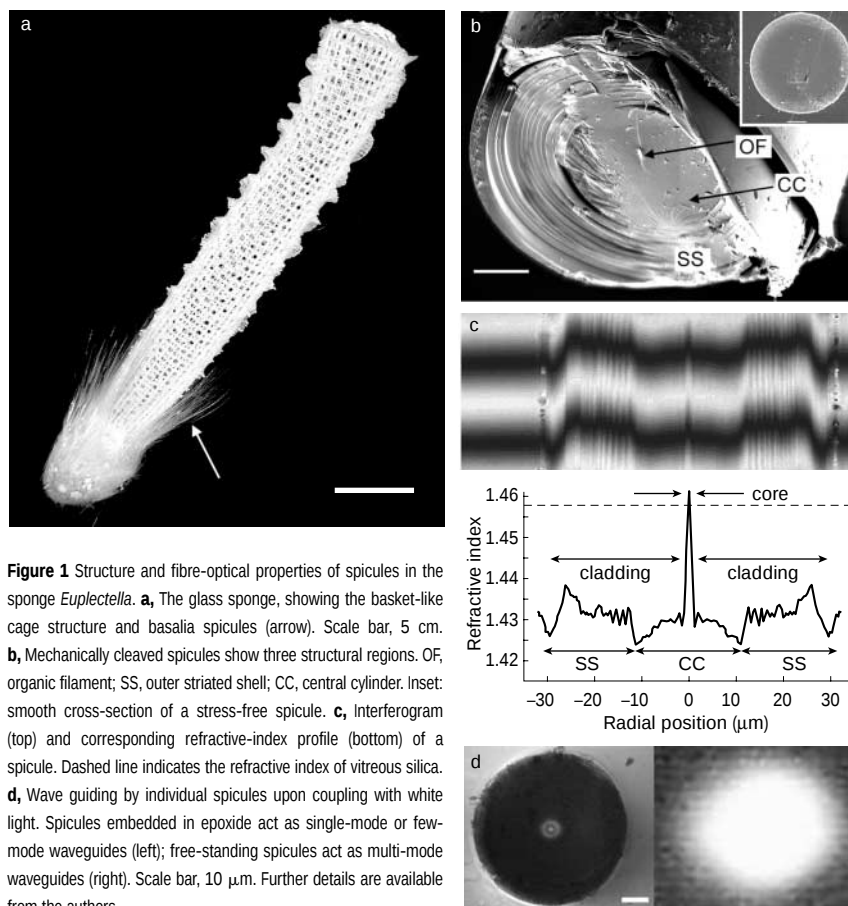


Figure 1 Structure and fibre-optical properties of spicules in the sponge *Euplectella*. **a**, The glass sponge, showing the basket-like cage structure and basalia spicules (arrow). Scale bar, 5 cm. **b**, Mechanically cleaved spicules show three structural regions. OF, organic filament; SS, outer striated shell; CC, central cylinder. Inset: smooth cross-section of a stress-free spicule. **c**, Interferogram (top) and corresponding refractive-index profile (bottom) of a spicule. Dashed line indicates the refractive index of vitreous silica. **d**, Wave guiding by individual spicules upon coupling with white light. Spicules embedded in epoxide act as single-mode or few-mode waveguides (left); free-standing spicules act as multi-mode waveguides (right). Scale bar, 10 μm . Further details are available from the authors.

1. Aizenberg, J., Tkachenko, A., Weiner, S., Addadi, L. & Hendler, G. *Nature* **412**, 819–822 (2001).
2. McPhedran, R. C. et al. *Aust. J. Chem.* **54**, 241–244 (2001).
3. Cattaneo-Vietti, R. et al. *Nature* **383**, 397–398 (1996).
4. Perry, C. C. & Keeling-Tucker, T. J. *Biol. Inorg. Chem.* **5**, 537–550 (2000).
5. Simpson, T. L. *The Cell Biology of Sponges* (Springer, New York, 1984).
6. Sarikaya, M. et al. *J. Mater. Res.* **16**, 1420–1428 (2001).
7. Marcuse, D. *Principles of Optical Fiber Measurements* (Academic, New York, 1981).
8. Clegg, W. J., Kendall, K., Alford, N. M., Button, T. W. & Birchall, J. D. *Nature* **347**, 455–457 (1990).
9. Kamat, S., Su, X., Ballarini, R. & Heuer, A. H. *Nature* **405**, 1036–1040 (2000).
10. Cha, J. N., Stucky, G. D., Morse, D. E. & Deming, T. J. *Nature* **403**, 289–292 (2000).
11. Kroger, N., Lorenz, S., Brunner, E. & Sumper, M. *Science* **298**, 584–586 (2002).

Competing financial interests: declared none.

Linguistics

Modelling the dynamics of language death

Thousands of the world's languages are vanishing at an alarming rate, with 90% of them being expected to disappear with the current generation¹. Here we develop a simple model of language competition that explains historical data on the decline of Welsh, Scottish Gaelic, Quechua (the most common surviving indigenous language in the Americas) and other endangered languages. A linguistic parameter that quantifies the threat of language extinction can be derived from the model and may be useful in the design and evaluation of language-preservation programmes.

Previous models of language dynamics have focused on the transmission and evolution of syntax, grammar or other structural properties of a language itself^{2–7}. In contrast, the model we describe here idealizes languages as fixed, and as competing with each other for speakers. For simplicity, we also assume a highly connected population, with no spatial or social structure, in which all speakers are monolingual.

Consider a system of two competing languages, X and Y, in which the attractiveness of a language increases with both its number of speakers and its perceived status⁸ (a parameter that reflects the social or economic opportunities afforded to its speakers). Suppose an individual converts from Y to X with a probability, per unit of time, of $P_{yx}(x, s)$, where x is the fraction of the population speaking X, and $0 \leq s \leq 1$ is a measure of X's relative status. A minimal model for language change is therefore

$$\frac{dx}{dt} = yP_{yx}(x, s) - xP_{xy}(x, s) \quad (1)$$

where $y = 1 - x$ is the complementary fraction of the population speaking Y at time t . By symmetry, interchanging languages

should yield the same transition probability as a swap in the fraction of speakers and relative status; thus $P_{xy}(x, s) = P_{yx}(1 - x, 1 - s)$. We also assume that no one will adopt a language that has no speakers ($P_{yx}(0, s) = 0$) or no status ($P_{yx}(x, 0) = 0$), and that P_{yx} is smooth and monotonically increasing in both arguments.

These mild assumptions imply that equation (1) generically has three fixed points. Of these, only $x = 0$ and $x = 1$ are stable. The model therefore predicts that two languages cannot coexist stably — one will eventually drive the other to extinction.

To test our model, we collected data on the number of speakers of endangered languages in 42 regions of Peru, Scotland, Wales, Bolivia, Ireland and Alsace-Lorraine, four instances of which are shown in Fig. 1. We fit the model's solutions to the data, assuming transition functions of the forms $P_{yx}(x, s) = cx^a s$ and $P_{xy}(x, s) = c(1 - x)^a(1 - s)$. Unexpectedly, the exponent a was found to be roughly constant across cultures, with $a = 1.31 \pm 0.25$ (mean $\pm$ standard deviation; further details are available from the authors).

Of the remaining parameters, status, s , is the most relevant linguistically; it could serve as a useful measure of the threat to a given language. Quechua, for example, still has many speakers in Huanuco, Peru, but its low status is driving a rapid shift to Spanish, which leads to an unfortunate situation in which a child cannot communicate with his or her grandparents.

Contrary to the model's stark prediction, bilingual societies do, in fact, exist. But the

histories of countries where two languages coexist today generally involve split populations that lived without significant interaction, effectively in separate, monolingual societies. Only recently have these communities begun to mix, allowing language competition to begin.

So what can be done to prevent the rapid disintegration of our world's linguistic heritage? The example of Quebec French demonstrates that language decline can be slowed by strategies such as policy-making, education and advertising, in essence increasing an endangered language's status. An extension to equation (1) that incorporates such control on s through active feedback does indeed show stabilization of a bilingual fixed point.

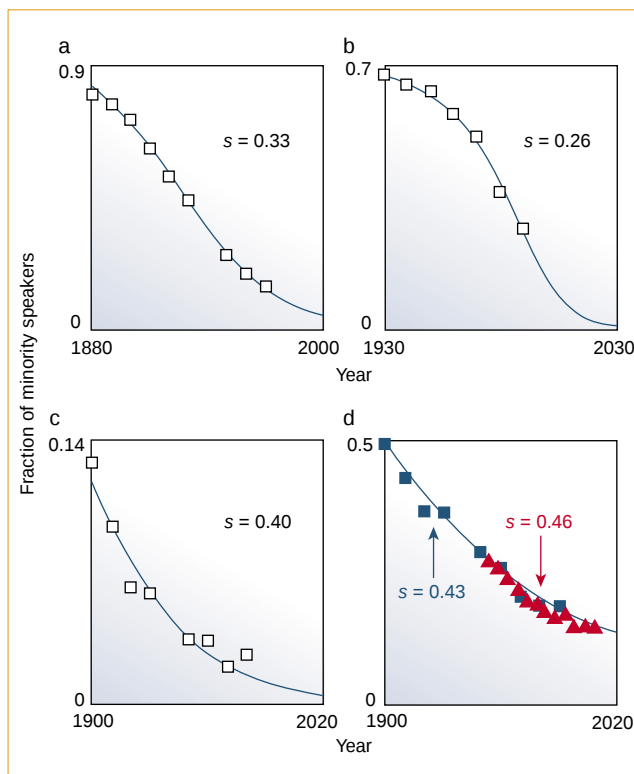
Daniel M. Abrams, Steven H. Strogatz

Department of Theoretical and Applied Mechanics, Cornell University, Ithaca, New York 14853, USA
e-mail: dma32@cornell.edu

1. Krauss, M. *Language* **68**, 4–10 (1992).
2. Nowak, M. A., Komarova, N. L. & Niyogi, P. *Nature* **417**, 611–617 (2002).
3. Chomsky, N. A. *Syntactic Structures* (Mouton, New York, 1957).
4. Kroch, A. *Lang. Variat. Change* **1**, 199–244 (1989).
5. Hawkins, J. A. & Gell-Mann, M. *The Evolution of Human Languages* (Addison-Wesley, Reading, Massachusetts, 1992).
6. Niyogi, P. & Berwick, R. C. *Ling. Phil.* **20**, 697–719 (1997).
7. Lightfoot, D. *The Development of Language: Acquisition, Changes and Evolution* (Blackwell, Oxford, 1999).
8. Fishman, J. A. *Reversing Language Shift* (Multilingual Matters, Philadelphia, 1991).
9. Withers, C. W. J. *Gaelic in Scotland 1698–1981: The Geographical History of a Language* (Donald, Edinburgh, 1984).
10. Aitchison, J. & Carter, H. A. *Geography of the Welsh Language* (Univ. Wales Press, Cardiff, 1994).
11. UK Office of Population Censuses and Surveys. *1991 Census: Report for Wales, 1* (HMSO, London, 1993).

Competing financial interests: declared none.

Figure 1 The dynamics of language death. Symbols show the proportions of speakers over time of: **a**, Scottish Gaelic in Sutherland, Scotland⁹; **b**, Quechua in Huanuco, Peru; **c**, Welsh in Monmouthshire, Wales¹⁰; **d**, Welsh in all of Wales, from historical data¹⁰ (blue) and a single modern census¹¹ (red). Fitted curves show solutions of the model in equation (1), with parameters c , s , a and $x(0)$ estimated by least absolute-values regression. Where possible, data were obtained from several population censuses collected over a long timespan; otherwise, a single recent census with age-structured data was used (although errors are introduced, the size of which are reflected in the differing fits in **d**). Using the fraction of Catholic masses offered in Quechua in Peru as an indicator, we reconstructed an approximate history of the language's decline.



The role of stomata in sensing and driving environmental change

Alistair M. Hetherington¹ & F. Ian Woodward²

¹Department of Biological Sciences, The Lancaster Environment Centre, University of Lancaster, Lancaster LA1 4YQ, UK

²Department of Animal & Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK

Stomata, the small pores on the surfaces of leaves and stalks, regulate the flow of gases in and out of leaves and thus plants as a whole. They adapt to local and global changes on all timescales from minutes to millennia. Recent data from diverse fields are establishing their central importance to plant physiology, evolution and global ecology. Stomatal morphology, distribution and behaviour respond to a spectrum of signals, from intracellular signalling to global climatic change. Such concerted adaptation results from a web of control systems, reminiscent of a 'scale-free' network, whose untangling requires integrated approaches beyond those currently used.

Stomata (Fig. 1) are small pores on the surfaces of leaves and stems, bounded by a pair of guard cells, that control the exchange of gases—most importantly water vapour and CO₂—between the interior of the leaf and the atmosphere. In this capacity they make major contributions to the ability of the plant to control its water relations and to gain carbon. Gas exchange is regulated by controlling the aperture of the stomatal pore and the number of stomata that form on the epidermis. Environmental signals such as light intensity, the concentration of atmospheric carbon dioxide and endogenous plant hormones control stomatal aperture and development. The acquisition of stomata and an impervious leaf cuticle are considered to be key elements in the evolution of advanced terrestrial plants¹, allowing the plant to inhabit a range of different, often fluctuating environments but still control water content. Here, we describe how the application of knowledge from cognate disciplines is providing new insights into how stomata evolve and are able to process information from simultaneous, often conflicting and sometimes rapidly changing signals. Although it is too early to say whether these recent advances will result in paradigm shifts in our understanding of how plants both respond to and drive environmental change, it is quite clear that stomata are a key experimental tool to investigate these phenomena.

Before considering specific aspects of stomatal biology it is important to reflect on the impact of stomata at the global level. Although the total stomatal pore area may be only 5% of a leaf surface², the rate of water vapour loss may reach as high as 70% of a

similar structure without a cuticle. Stomata exert major controls on both the water and carbon cycles of the world. Annual precipitation over the land is about 110,000 km³, or 110×10^{15} kg (ref. 3) and evaporation and transpiration total about 70×10^{15} kg. The contribution of stomatal transpiration alone to the global water cycle can be determined by using a dynamic vegetation model⁴ (Fig. 2). The greatest rates of transpiration occur in the uniform and warm forested areas between the tropics with 32×10^{15} kg yr⁻¹ of water vapour passing through stomata. This is double the water vapour content of the atmosphere (15×10^{15} kg yr⁻¹). Terrestrial gross photosynthesis annually fixes about 120×10^{15} g C (440×10^{15} g CO₂) from the atmosphere's 730×10^{15} g C (ref. 5). The global distribution of this flux parallels the distribution of transpiration, indicating the closely coupled controls of stomata on CO₂ and water vapour diffusion.

Stomatal evolution

Stomatal control of water loss allows plants to occupy habitats with fluctuating environmental conditions and so it can be predicted that stomata must be important contributors to speciation and evolutionary change. Stomata first appeared in terrestrial land plants over 400 million years ago (Myr)⁶ and since then have changed markedly in size and density on plant surfaces. There are two broad morphological types of stomata, the dumb-bell-shaped stomata typical of the grasses and the kidney-shaped form found in other species (Fig. 1).

Is there any evidence that stomata are involved in speciation?

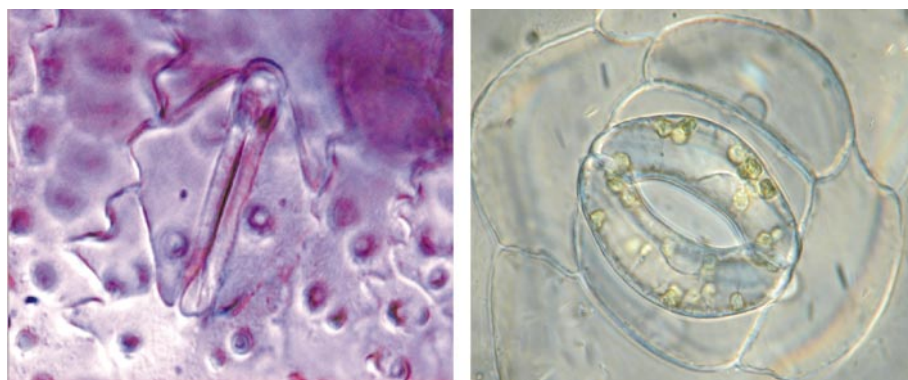


Figure 1 Dumb-bell-shaped stoma of rice typical of the grasses (left) and the kidney-shaped stoma typical of other species such as *Arabidopsis* and *Commelina* (right).

Intriguingly, over the 400 million years of the Phanerozoic era, periods of low atmospheric CO₂ concentrations are associated not only with high stomatal densities⁷ but also the emergence of new plant groups such as the ferns, pteridosperms and angiosperms⁸. Whether this involves some interaction and indeed causation requires further investigation. Correlations between changes in global environmental conditions and stomatal evolution can be demonstrated for species in the Proteaceae and for the evolution of dumb-bell stomata in the Poaceae. These changes occurred in the Cenozoic era of the last 66 million years when there were profound changes in global climate and terrestrial flora and fauna⁹.

Chloroplast DNA has been used to derive a phylogeny for species of *Banksia* and *Dryandra* in the Proteaceae of Australia¹⁰. Two clades of species are currently recognized by differences in stomatal distributions. In the clade Cryptostomata, stomata occur in shallow pits or in crypts, whereas they have a more superficial distribution in the clade Phanerostomata. The superficial occurrence appears to be the primitive state and species in the clade Phanerostomata occur in moist climates (such as the most recent common ancestor). Species of the clade Cryptostomata occur in much drier climates and probably diverged from the Phanerostomata clade 55–35 Myr, at a time when the climate was becoming more arid¹¹. Although stomatal differences are not the only differences between the clades, the marked differences in stomatal location would have exerted differential capabilities for the spread and survival of the two clades in moist and arid climates¹¹. The environmental correlates of the differences in stomatal distribution seen for the Proteaceae are nicely demonstrated in *Cistus incanus*, for which similar differences in stomatal distribution occur, but between the summer and winter of a Mediterranean climate¹². Leaves produced in the cool and wet winter are large and flat with frequent stomata on the abaxial leaf surface; however, leaves developed in the hot and dry summer are crimped and partially rolled, forming a crypt on the lower surface, the only location of stomata.

The Poaceae consists of about 10,000 species for which the macro-evolutionary history has recently become established by the analysis of chloroplast and nuclear DNA¹³. The linear dumb-bell-shaped stomata of grasses (Fig. 1) are generally believed to represent a more evolutionary advanced form than their kidney-shaped counterparts. This is supported by the observation¹⁴ that

during development, Timothy grass guard cells adopt a transient kidney-shaped phase before assuming their typical (mature) dumb-bell shape. The linear dumb-bell design magnifies small changes in width to cause large openings, and maximizes the potential of the stomata to track changes in environmental conditions, probably with little energetic cost. Smaller changes in guard and subsidiary cell turgor lead to greater increases in stomatal aperture¹⁵ in the dumb-bell-shaped stomata than occur for kidney-shaped stomata. This efficiency and speed of stomatal opening in grasses enhances photosynthesis and water use efficiency compared with non-grass species¹⁶. A rapid stomatal response to blue light augments photosynthesis in early morning and in intermittent sunlight, in which light has an enhanced blue light content¹⁶ and which would have characterized the understorey environment during the early evolution of grasses. The low aerodynamic conductance of a grassland canopy could reduce the impact of changes in stomatal aperture on gas exchange⁸. However, field observations¹⁷ demonstrate a limited impact of aerodynamic conductance on stomatal dynamics.

Grasses originated between about 55 and 70 Myr (ref. 13), leading to lineages that were understorey plants of tropical forests. Their spread and diversification, during global aridification 30–45 Myr, would have been enhanced by the dumb-bell stoma, capable of responding quickly and efficiently to the enhanced light conditions of newly open habitats, but with the capacity to avoid the increased likelihood of drought. This period just preceded the diversification of grazing animals^{9,18} and was well before the origin of the C₄ pathway of photosynthesis. Animal grazing and browsing of grazing-intolerant shrubs and trees would have enhanced the spread of grazing-tolerant grasses, particularly into areas of open woodland, whereas the higher albedo of the grasslands may have enhanced regional aridity¹⁸.

Environmental control of stomatal development

We shall not discuss here the details of stomatal development but rather the control of stomatal distribution and size by environmental factors. Depending on the species and the environmental conditions stomata range in size from about 10 to 80 µm in length and occur at densities between 5 and 1,000 mm⁻² of epidermis² (Fig. 3a). In spite of this wide variability there is a strong and general relationship between density and size (Fig. 3a) for different plant

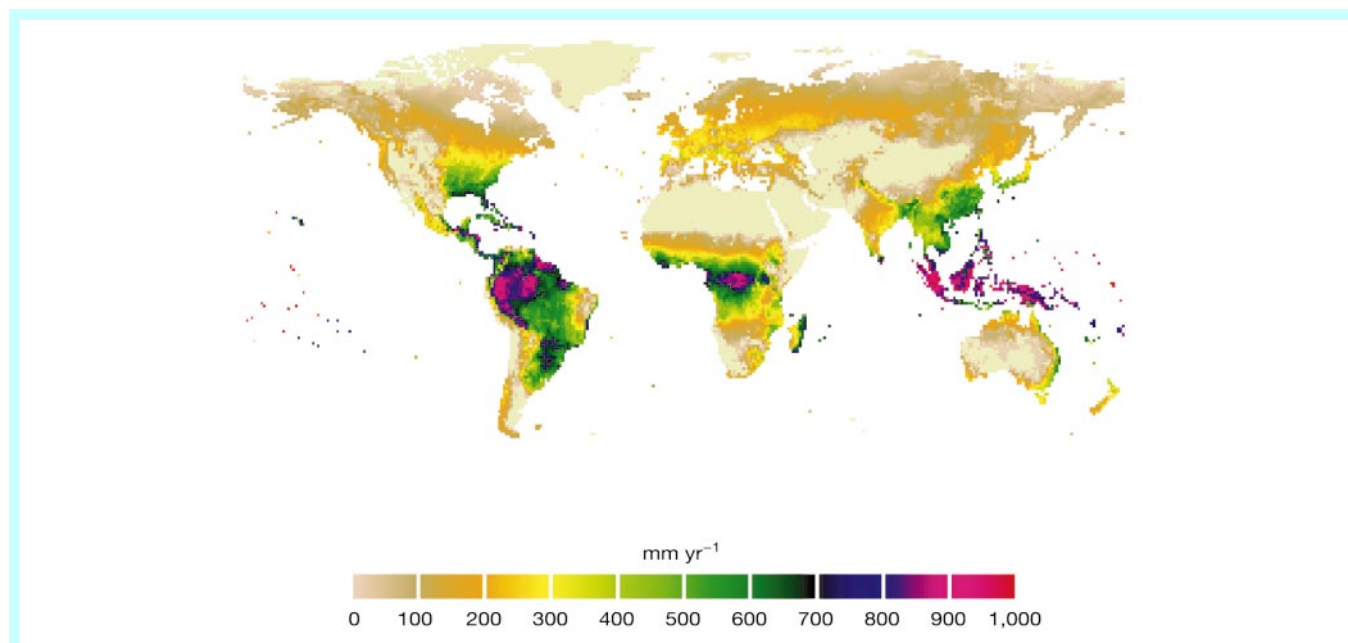


Figure 2 Transpiration (mm yr⁻¹) from terrestrial vegetation simulated with the Sheffield Dynamic Global Vegetation Model (SDGVM)⁴ and averaged for the 1990s.

groups (grasses and non-grasses), fossil leaves and for different stomatal distributions on either one or both leaf surfaces. Simulations with a stomatal model of gaseous exchange⁷ indicate that the average relationship in Fig. 3a has virtually the same trend to one in which changes in density and size are exactly compensatory, in terms of CO₂ and water vapour exchange. Apart from the earliest land plants of the Silurian and Devonian periods, with very low stomatal densities and small sizes, it also seems that this relationship has existed for the last 300 million years. On this basis it is not obvious how selection favours any particular species' characteristic of stomatal density and size.

The plant must maintain movement of water from the soil to the leaf, and rapid stomatal responses to environmental change are a major feature of this maintenance¹. One study¹⁹ has demonstrated that stomatal size has a key role in this control and for six forest trees there is a clear negative relationship between the length of the stomatal pore and sensitivity to increasing drought. In these species larger stomata (species with kidney-shaped stomata) were slower to close and demonstrated a greater potential for hydraulic dysfunction under drought. Ferns from deep shade possess large stomata at low densities²⁰ and in this natural environment, which may be cool and humid, it is found that truly shade-tolerant species often retain

open stomata, even in deep shade, at least for early parts of the day²¹. The constancy of the open stomata will minimize the impact of what would otherwise be slow opening limitations to photosynthesis during short-lived periods of sunlight, which are critical for enhancing photosynthesis in this light-limited environment. Therefore the limited available information suggests that large kidney-shaped stomata seem to be an important feature for plants of humid and deep shade conditions but their slow dynamic behaviour could lead to problems under dry conditions. Small stomata can open and close more rapidly and their general association with high densities (Fig. 3a) provides the capacity for rapid increases in the stomatal conductance of a leaf, maximizing CO₂ diffusion into the leaf during favourable conditions for photosynthesis¹⁹.

The effect of growth at elevated concentrations of CO₂ on stomatal density and stomatal index (the fraction of epidermal cells that are stomata) is one of the most intensively studied environmental controls on stomatal development. CO₂ enrichment changes the stomatal density of different species and different accessions (ecotypes) of *Arabidopsis thaliana* (Fig. 3b and ref. 22). With stomatal densities ranging from 45 to 720 mm⁻² the mean response is an 11% reduction in density with a doubling of the CO₂ concentration (Fig. 3b), and which is insensitive to the basal

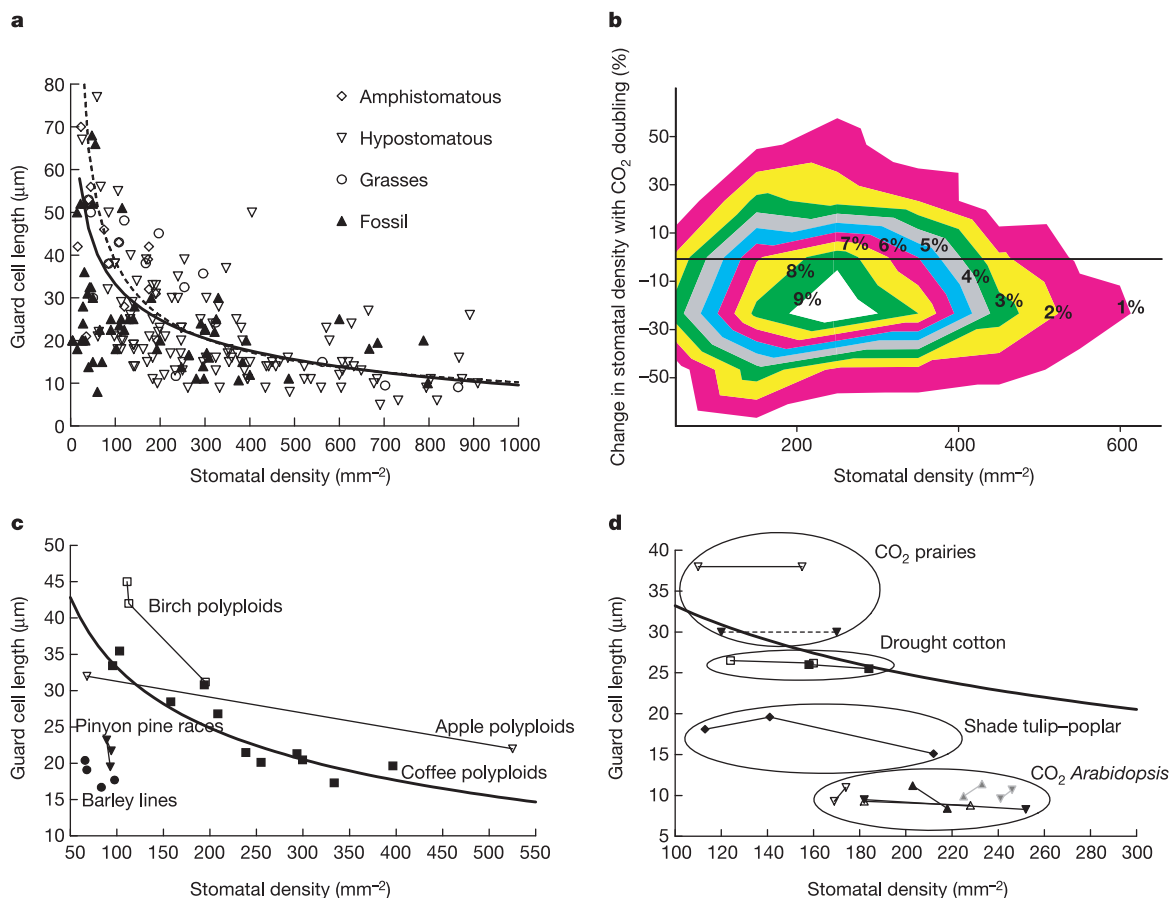


Figure 3 Control of stomata by the environment. **a**, Relationship between stomatal width and density. Data from amphistomatous species, hypostomatous species, grasses and fossil leaves are shown. The data are from refs 2, 7, 67–69 and F.I.W. (personal observations). The solid curve is log-normal, $y = -28.75 + 162x^{-0.2086}$, $r^2 = 0.5$; the dashed line shows equal stomatal conductance with variable stomatal density. **b**, Responses (per cent change) of stomatal density to CO₂ doubling for 125 species and 63 accessions of *A. thaliana* (from ref. 22 and new observations). Smoothed contours indicate fraction of genotypes (%) with particular values of stomatal density and per cent response to enrichment. **c**, Variations in stomatal characteristics

with different genotypes. Curve as for **a**. Birch polyploids ($2n, 5n, 6n$)⁷⁰, apple polyploids ($3n, 4n$)⁷¹, different races of pinyon pine⁷², drought-selected lines in barley⁷³ and different polyploids in coffee⁷⁴ are shown. **d**, Stomatal responses to different environmental conditions. Curves as for **a**. CO₂ enrichment experiments, prairies⁷⁵; *Arabidopsis* (our own unpublished data) where adaxial surface (filled triangles) and abaxial surface (inverted filled triangles) are indicated; tulip-poplar changes through a canopy⁷⁶; and drought responses of cotton⁷⁷ are shown. Different ecotypes are represented by different colours within groups.

stomatal density. The reduction in stomatal density with CO₂ enrichment leads generally to a decrease in maximum stomatal conductance but an increase in the maximum rate of photosynthesis, at the elevated CO₂ concentration^{7,8,22}.

Royer²³ established that the reductions of stomatal density and stomatal index associated with growth under increased concentrations of atmospheric CO₂ were recorded more frequently in observations from longer-term measurements of stomatal responses (herbarium material with a decade to century timescale

and from fossil leaves with a millennial timescale) than in short-term field or growth chamber seasonal experiments. Royer suggested that this increased frequency of response was due to the transition from a variable plastic response by species in short-term experiments to a genetic response, as a result of selection on longer timescales. Figure 3c supports the suggestion that genetic change can alter the stomatal density–size relationship (Fig. 3a), whereas environmental changes seem primarily to move the position of a species along the stomatal size–density curve (Fig. 3d). Changes in the degree of ploidy can also change stomatal density and size (Fig. 3c). Notably, in *Arabidopsis*, an unresponsive accession (C24) for the stomatal density response to CO₂ enrichment is actually regulated to show this minimum response because the mutation of a single gene leads to a very significant response to CO₂ (ref. 24).

Currently we know rather little about the signalling pathways by which environmental signals control stomatal development. Recent work has shown that in *Arabidopsis* the *HIC* gene, which encodes a putative 3-ketoacyl CoA synthase (KCS), is involved in the control of stomatal development by elevated concentrations of CO₂ (ref. 24). As KCS is involved in the synthesis of wax components found in the cuticle it has been suggested that in *hic* an alteration to cuticular structure and properties interferes with the diffusion of an endogenous inhibitor that controls stomatal development²⁴. This suggestion receives support from the observation that some *Arabidopsis* wax-deficient mutants display abnormal stomatal patterning^{24,25}. Analyses of other *Arabidopsis* mutants²⁶ indicate that the CO₂ response of stomatal index is absent in *fad-4*, a jasmonic acid mutant, whereas in *ein-2*, an ethylene-insensitive mutant, the CO₂ response is absent only from the adaxial leaf surface. The jasmonate and ethylene transduction pathways are also involved in defence responses to pathogens²⁷ and the *ein-2* mutant is susceptible to attack by pathogens. It is remarkable that mechanisms for addressing pathogen attack are also central in the responses of stomatal development to CO₂ concentration. Accessions of *Arabidopsis* differ widely in resistance both to powdery mildew disease²⁸ and to CO₂ concentration²², and it is tempting to link these two major responses, but experimental support is still wanting. Any connection between the two processes would influence strongly the processes of selection for the stomatal response to CO₂, in the long term, supporting the notion that the response is more reliable the longer the period of study²³. More recent work²⁹ shows that, similar to responses to pathogen attack³⁰ there is also systemic control of stomatal development during growth under elevated CO₂ as mature leaves both detect CO₂ and produce a signal to influence development of younger expanding leaves.

The work on the *ein-2* mutant of *Arabidopsis*²⁶ suggests that ethylene differentially controls stomatal development on the upper and lower leaf surfaces. Comparing stomatal densities among *Solanum pennellii*, *Lycopersicon esculentum* and a graft-induced periclinal chimaera, having an *S. pennellii* epidermis and an *L. esculentum* mesophyll, showed³¹ that the two surfaces could behave independently. Both donors had more stomata on the lower than on the upper surface and *L. esculentum* had a greater stomatal density than *S. pennellii*. The stomatal density on the upper epidermis of the chimaera was similar to that of the epidermal donor *S. pennellii*. However, the density on the chimaeral abaxial epidermis was significantly lower than its donor. These data suggest that stomatal differentiation is subject to different controls on each surface, and at least in the case of the lower epidermis a role for the mesophyll seems possible.

Control of stomatal aperture by environmental signals

Recent work shows that the control of stomatal aperture by environmental signals depends on coordinated alterations to guard cell turgor (ionic fluxes and sugars), cytoskeleton organization, membrane transport and gene expression (see refs 32 and 33 and references therein). A number of lines of evidence suggest that

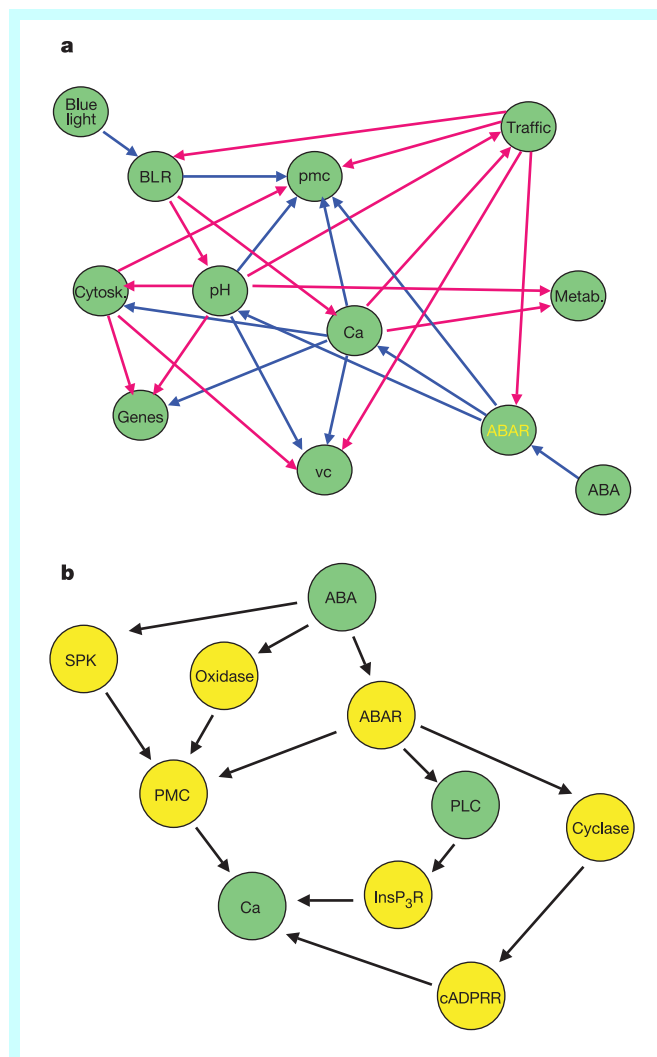


Figure 4 Model of guard cell signalling. **a**, Guard cell ABA and blue light signalling represented by a network-based model. In this model the nodes represent modules or groups of functionally related processes or second messengers. Modules linked by blue connections indicate that these links exist in guard cells. Red connections indicate that these links occur in other plant cells but their existence has not been investigated in guard cells. BLR, blue light receptor; pmc, plasma membrane ion channels; traffic, membrane trafficking; cytosk., cytoskeleton; vc, vacuolar ion channels; metab., metabolism of starch, sucrose, malate; ABAR, ABA receptor. Data are from refs 32, 33, 42. **b**, Guard cell signalling is robust. Multiple routes for generating increases in cytosolic Ca²⁺ concentration. Intracellular messengers (InsP₃, cADPR) and local messengers (sphingosine-1-phosphate, H₂O₂) are not shown. Components shaded in green have been isolated and characterized in guard cells. Other components (yellow) are predicted to exist on the basis of physiological or pharmacological evidence. For simplicity, possible feedback links are not shown. SPK, sphingosine kinase; oxidase, NADPH oxidase activity; PMc, plasma membrane calcium-permeable channels; ABAR, ABA receptor; PLC, phospholipase C; cyclase, ADP ribosyl cyclase; cADPRR, cyclic ADP ribose receptor; InsP₃R, InsP₃ receptor. Data are from refs 32, 33.

stand-alone, stimulus-specific signalling pathways might be an inadequate means of controlling stomatal aperture. First, and most importantly, the recent recognition that the control of stomatal aperture requires the coordinated control of multiple cellular processes must at the very least require repeated pathway bifurcation. Factoring in the requirement for feedback and temporal coordination of these different processes indicates that the architecture must become increasingly reticulate. In addition the presence of intracellular signalling components that feature in multiple signal pathways suggests that the existence of truly stand-alone pathways are highly unlikely. A good example of this phenomenon³⁴ is cytosolic free calcium ($[Ca^{2+}]_{\text{cyt}}$), others are the *Arabidopsis* *ABI1* and *ABI2* genes that encode type PP2C serine threonine phosphatases³². These phosphatases are involved in guard cell abscisic acid (ABA)³⁵, $[CO_2]$ ^{36–38} and (lack of light) darkness³⁸ signalling. It is difficult to imagine how a signalling component could be isolated to the extent that it only affected a single pathway. The presence of common signalling components argues strongly for interaction between pathways. Although these potential problems do not rule out the possibility that guard cells contain stand-alone signalling pathways, they do provide the justification for examining whether there are alternative organizational structures. Is there any evidence that guard cell signalling is organized as a network and specifically as a type of network known as a scale-free network?

Guard cell signalling: an example of a scale-free network?

Scale-free networks are robust and flexible and are ideally suited for discriminating and processing multiple signals simultaneously. These properties arise from the network itself rather than from individual components^{39–41}. Scale-free networks are composed of many interconnected nodes. In the context of this review, nodes equate to signalling components. The network contains a small number of highly connected nodes, known as hubs, which are central to the functioning of the scale-free network. Scale-free networks are characterized by a power law distribution of nodes (many sparsely and few highly connected nodes) and exhibit emergent properties of robustness against errors but fragility against the removal or damage to the hubs³⁹. Defining whether a network is scale-free requires a large network map of at least 1,000 nodes, which is analysed for a power law distribution. Unfortunately, this analysis is not possible in guard cells because the current guard cell signalling map is both too small and incomplete. However, it is possible to arrange the elements in guard cell signalling as a network (Fig. 4a). It is also possible to ask whether guard cell signalling displays properties similar to the emergent properties of scale-free networks. Even though, as we shall see, guard cell signalling does exhibit properties reminiscent of scale-free networks, it must be stressed that to define the guard cell signalling system as scale-free with any rigour requires that the distribution of nodes must be measured and described by a power law.

Guard cell signalling is robust

Scale-free networks are robust, which means that they exhibit a high degree of tolerance to node removal. There are data showing that guard cells continue to function despite the loss of certain signalling components. For example *Arabidopsis* plants carrying mutations in either of the blue light receptor genes *PHOT1* or *PHOT2* exhibit wild-type blue-light-induced stomatal opening. Only when both receptor genes in the *phot1*, *phot2* double mutant are disrupted do the stomata fail to respond to blue light⁴². A second example of robust behaviour comes from ABA signalling. ABA can generate increases in $[Ca^{2+}]_{\text{cyt}}$ using plasma membrane calcium-permeable channels, the PI-PLC/inositol-1,4,5-trisphosphate (InsP₃) pathway, cyclic ADPribose, sphingosine-1-phosphate and possibly inositol hexakisphosphate (reviewed in ref. 40; see also Fig. 4b). Recent studies using tobacco, in which the levels of PI-PLC protein have been reduced in guard cells, reveal a very modest effect of interfering

with this signalling system on the ability of stomata to respond to ABA⁴³. It is possible to conclude from these investigations that when PI-PLC levels are reduced, the guard cell is able to compensate for this loss by making use of other calcium mobilizing systems or other PI-PLCs. These are two examples in which guard cells exhibit behaviour reminiscent of the robust properties of a scale-free network.

What advantage might guard cell robustness confer on the plant? One possibility is that it will protect, to some extent, the vital process of carbon acquisition and the regulation of water loss from the (possibly) deleterious effects of mutation in genes encoding guard cell signalling components. In addition the presence of multiple copies of components or mechanisms for generating a similar outcome provides the raw material on which evolution could operate.

Guard cell signalling is fragile to hub removal

Fragility to hub removal is a well-characterized property of scale-free networks³⁹. Does guard cell signalling exhibit similar behaviour? In the signalling context a hub can have a central role in coupling stimulus responses. On the basis of our current knowledge of guard cell signalling perhaps the best candidate for a hub is the increase in the concentration of guard cell $[Ca^{2+}]_{\text{cyt}}$ (Fig. 4b) that occurs in ABA, extracellular calcium ion, hydrogen peroxide, CO_2 and IAA signalling^{32,33}. As mentioned above, there are multiple mechanisms for generating an increase in guard cell $[Ca^{2+}]_{\text{cyt}}$. Interfering with one of these results in very small phenotypic effects, which is reminiscent of the robust properties of scale-free networks. However, when the increase in $[Ca^{2+}]_{\text{cyt}}$ is totally prevented by microinjection of the calcium chelator BAPTA into the guard cell cytosol there is no ABA-mediated loss of turgor⁴⁴. This is an example where interfering with the activity of a hub results in the failure of the system. Taken together, these results show that guard cell signalling exhibits similar properties to scale-free networks, however, as pointed out above, testing whether this is an accurate description of the architecture of the guard cell signalling system requires much more data and a rigorous mathematical analysis.

Acclimation and rhythmic behaviour in guard cells

If guard cell signalling is organized as a network then a striking property of the network is that it acclimates to external signals. We are not aware whether acclimation is an emergent property of scale-free networks. Stomata of *Vicia faba* grown in a growth chamber are markedly more sensitive to CO_2 than stomata from plants grown in a greenhouse⁴⁵. These alterations in CO_2 sensitivity were also observed for stomata present in isolated epidermal preparations, suggesting that guard cell behaviour, rather than the properties of the leaf as a whole, had been modified.

The response of stomatal conductance to CO_2 doubling is highly variable but observations on trees, lasting up to 4 yr (ref. 46), demonstrated an average reduction of 21%. Some of the response may also include changes in stomatal density⁴⁷, although this was not measured directly.

Although we do not yet know the cellular basis of acclimation in guard cells the issue does have a bearing on the organization of guard cell signalling components. If CO_2 signalling were a stand-alone pathway we would predict that acclimation would have no effect on the response of guard cells to other signals. However, if the effects of CO_2 on stomatal aperture are brought about through a signalling network then alterations in sensitivity to this signal should have effects on other pathways. Is this the case? When *Arabidopsis* is grown under ambient concentrations of CO_2 and then exposed to elevated CO_2 , stomatal conductance decreases⁴⁸. This was transient and conductance returned to pre-treatment levels quickly, suggesting that stomata are able to acclimate rapidly to elevated CO_2 . Plants were then grown at ambient and elevated CO_2 for two days and stomatal conductance measured after exposure to

osmotic stress or treating them with ABA. Enhanced responses to both treatments were observed in the plants grown at elevated CO_2 relative to their ambient controls⁴⁸. These data suggest that acclimation to one signal leads to alterations in sensitivity to another signal and are consistent with a network-based organization. However, it should be noted that effects were not obvious, at least in the case of ABA applied to isolated epidermal strips.

Stomata are influenced by rhythms. These either control stomatal aperture directly or modulate the response of stomata to other signals. The most thoroughly investigated of these are the circadian rhythms that in C_3 and C_4 plants result in stomatal opening during the day and closure at night⁴⁹. The circadian clock can also modulate the sensitivity of stomata to exogenous signals such as light⁵⁰. Studies using *V. faba* revealed that stomatal opening at the start of the day was primarily associated with K^+ accumulation whereas later in the day sucrose was the dominant solute. Differential circadian regulation of the components responsible for the accumulation or loss of these solutes might underlie this phenomenon. There is also evidence, which would be worthy of re-investigation using contemporary controlled growth facilities, of an annual rhythm affecting stomatal apertures⁵¹. This annual variation in aperture is mirrored to some extent by the physiological properties of guard cells, where there is seasonal variation in membrane potential⁵² and plasma membrane (H^+)ATPase activity⁵³.

Autonomous units and collective behaviour

Early in development guard cells lose plasmodesmatal connections with each other and their neighbours⁵⁴. This symplastic isolation explains why injecting a single guard cell from a stoma with cyclic ADP ribose causes reductions in turgor in the injected cell while the uninjected partner remains unchanged⁵⁵. Although symplastic isolation makes it possible for an individual stoma to behave as an autonomous unit, is such behaviour likely to have physiological significance? A stoma is unlikely on its own to have a significant impact on leaf processes. However, groups of stomata may exhibit localized coordinated behaviour and this can benefit the plant. Indeed, this is what probably underlies the phenomenon of patchiness in which some areas of the leaf display relatively open stomata while other parts of the same leaf may exhibit reduced apertures⁵⁶. The benefits of localized semi-autonomous behaviour are most readily seen in plants with relatively large leaves that inhabit habitats characterized by localized unpredictable variations in environmental signals. Perhaps the best-investigated example is the response of leaves in the forest understorey to sunlight penetrating through the canopy and illuminating areas of the leaf in a highly localized

manner. The adaptive significance here is fairly obvious. Opening of the stomata in the illuminated region of the leaf, while maintaining the remaining stomata in a relatively closed state in the dim light of the understorey would allow the plant to exploit the available light efficiently without losing unnecessary amounts of water²¹. There will also be attendant energy savings to opening in restricted sectors of the leaf. In this context semi-autonomous behaviour can be seen to support the opportunistic capture of resources. However, it should be pointed out that localized collective behaviour has recently been interpreted as evidence for non-autonomous behaviour of stomata⁵⁷. Of course in other situations where the plant has small leaves or the leaf experiences homogeneous exposure to the dominant environmental variable the result will be uniformity of stomatal aperture. However, even emergent collective behaviour is still the result of signals being perceived by and operating on individual stomata.

Stomatal impact on photosynthesis and transpiration

The diffusion of CO_2 into the leaf during photosynthesis and the outward diffusion of water vapour are both controlled by the opening and closing of stomata, although the diffusion rate of water vapour is 1.6 times greater than CO_2 . There is also another measure of stomatal control, the relationship between the maximum conductance of stomata (determined by the opening and density of stomata) and the maximum rate of photosynthesis measured under optimal conditions (Fig. 5). Short-term field measurements demonstrate a wide range of rates of photosynthesis (A) and stomatal conductance (g). However, leaves of different species and life forms fall, with some scatter, on either of two curvilinear relationships. Plants with the C_4 pathway of photosynthesis show greater rates of photosynthesis than those with the C_3 pathway. There is also a close to constant ratio between the substomatal or intercellular CO_2 concentration and the CO_2 concentration in the ambient air. The ratio falls between about 0.65 and 0.8 for C_3 species and between 0.4 and 0.6 for C_4 species. This indicates a close matching of the diffusional supply of CO_2 to the chloroplasts of the leaf and the demand from the photosynthetic enzymes, a central feature of the photosynthetic process⁵⁸. The mechanism that maintains this close relationship is still unclear, and reflects the fact that unlike photosynthesis, a fully mechanistic model of stomatal activity does not exist⁵⁹.

The slope of the relationship between A and g ($\text{d}A/\text{d}g$) (Fig. 5) is shown for a C_3 species and is a measure of water use efficiency (the rate of carbon gain per unit increase in stomatal conductance). The greatest efficiency occurs at low stomatal conductances but flattens quickly at conductances above about $0.4 \text{ mol m}^{-2} \text{ s}^{-1}$. The relationship between transpiration, E , and conductance ($\text{d}E/\text{d}g$) has the same form as $\text{d}A/\text{d}g$, whereas the response of leaf temperature, T ($\text{d}T/\text{d}g$) is inversely related to $\text{d}E/\text{d}g$. This means that stomata can exert significant control over both leaf temperature and gaseous diffusion. Small changes in g exert the greatest impacts on leaf temperature when g is small, with a flattening of response above $0.4 \text{ mol m}^{-2} \text{ s}^{-1}$. Plants with higher stomatal conductances show minimal gain of photosynthesis and minimal changes in leaf temperature for large changes in conductance, reflecting high water usage and little limitation on water supply. By contrast plants with low values of g may occur in water-limited environments, where small reductions in stomatal conductance significantly increase water use efficiency. The continued increase in atmospheric CO_2 concentration will enhance photosynthesis, at least in C_3 species, and decrease stomatal conductance⁴⁶ and so plants will become more efficient at water use and may use less water⁶⁰. However, species differ significantly in these CO_2 responses, with some species being more favoured than others, indicating that differing species-specific responses of stomatal conductance and photosynthesis to CO_2 have the capacity to change community composition⁶¹.

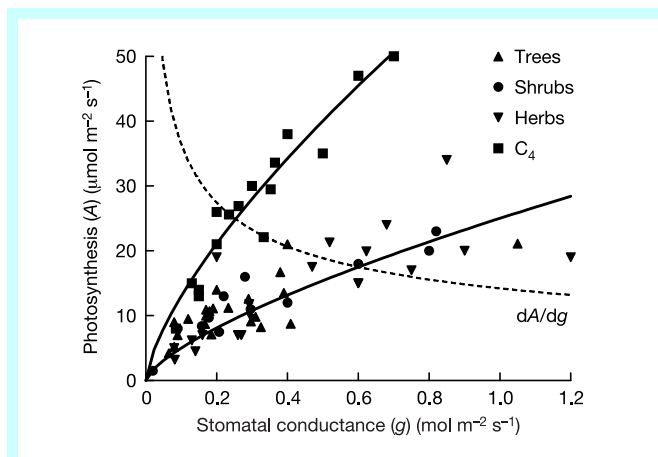


Figure 5 Field observations of maximum photosynthesis (A) and stomatal conductance to water vapour (g). Data are from refs 78–100. The dotted line is the differential of the line of best fit between A and g for the C_3 species, and indicates $\text{d}A/\text{d}g$ ($\mu\text{mol mol}^{-1}$).

Field experiments have demonstrated that differences in water supply can select for a wider range of plant responses than suggested from Fig. 5. About half of the plants with stomatal conductances less than $0.4 \text{ mol m}^{-2} \text{ s}^{-1}$ occur in dry conditions, whereas about 40% of those with conductances greater than $0.4 \text{ mol m}^{-2} \text{ s}^{-1}$ are also found in dry habitats. Plants with high g , in dry conditions, demonstrate opportunistic and rapid growth during short periods of water availability, in which high conductances and photosynthetic capacity will maximize growth. This contrasts with other plants from dry conditions with less active periods of growth, and low values of g , but with a more extended period of activity.

In cotton and wheat, yields are positively correlated with maximum stomatal conductance, but not with photosynthesis⁶². In this case high stomatal conductances lead to significant evaporative cooling of irrigated crops, to temperatures below the threshold for yield reduction. Experiments on *Cakile edentula* growing in cooler conditions demonstrated that fitness in a dry environment was greatest for individuals with the highest water use efficiency⁶³, indicating selection at the leaf level on the balance of carbon uptake and water loss. In the dry environment stomatal conductance was negatively correlated with fitness, whereas water use efficiency was not selected in a wet environment. An example of differential selection in nature for stomatal behaviour with water supply has been demonstrated for hermaphrodite and dioecious populations of the species *Wurmbea dioica*⁶⁴. These populations occur sympatrically, but the hermaphrodite populations occur in wetter microsites. Even though they are found at wetter sites, the hermaphrodite populations adjust stomatal conductance to minimize water loss, which allows greater photosynthetic allocation to leaves and reproductive structures. The dioecious populations, from the drier sites, maximize CO_2 uptake and water loss, with maximal stomatal conductances, in parallel with increased allocation to storage organs.

Future prospects

In this review we have advanced the hypothesis that guard cell signalling is organized as a scale-free network. Before we can rigorously test whether this is an accurate description we need to identify more components that contribute to guard cell signalling. An increase in the application of forward and reverse genetic strategies to guard cell function and development combined with protein–protein interaction studies are likely to produce the data sets required to test whether the power law distribution applies to guard cell signalling components. It is worth pointing out that if a network or more specifically, as we suggest here, a scale-free network does reflect the organization of the signalling components in the guard cell rather than a collection of linear pathways, we shall need to re-evaluate how we interpret the ‘positioning’ of signalling components relative to one another using epistasis analysis. Avery and Wasserman⁶⁵ identified a set of conditions that need to be satisfied before it is possible to draw meaningful conclusions from epistasis analyses. One of these is that “the signal and two genes under study are the sole determinants of the phenotype under the conditions of the experiment”. Scale-free networks with their inherent redundancy and robustness would seem to contravene this rule. Similarly, the application of this rule would make the interpretation of data gained by using stomatal aperture as the phenotype problematic, as this is clearly controlled by multiple signals acting simultaneously. Accordingly, until we know more about the architecture of the guard cell signalling system, we suggest that the positioning of components relative to one another using epistasis analysis is treated with some caution. Finally, in the context of signalling it would be interesting to determine whether defined scale-free networks exhibit properties such as acclimation that are typical of guard cell signalling.

We also need to know much more about the underlying cell biology of the dumb-bell-shaped stomata of the grasses. The

evolution of the dumb-bell stomata may have been a major component not only in the geographical spread of grasses but also in the coevolution of grazing animals and regional climate. Geological evidence indicates a three-way interaction between the evolution of grazing tolerance in grasses, the rise of grazing ungulates and the decline in browsers since the end of the Eocene epoch, 38 million years ago¹⁸. The end of the Eocene epoch was characterized by a drier climate, which would have favoured the spread of drought-tolerant grasses into vegetation that had been dominated by trees and shrubs⁶⁶. In this respect the evolution of the efficient dumb-bell stomata of grasses and their impact on the ability of the grasses to dominate new habitats can be seen as a prerequisite for the evolution and subsequent domestication of grazing animals by man. Given the importance of grasses ecologically and in the evolutionary context and also as commercial crops we need to understand much more about their development and function. In the past efforts to investigate these subjects have been impeded by the tough lignified guard cell wall that resists protoplasting and impedes microinjection. With the completion of the rice genome it seems likely that we are set to gain many insights into the mysteries of grass stomatal development and evolution. □

doi:10.1038/nature01843.

1. Raven, J. Selection pressures on stomatal evolution. *New Phytol.* **153**, 371–386 (2002).
2. Willmer, C. & Fricker, M. *Stomata* 2nd edn (Chapman & Hall, London, 1996).
3. Jackson, R. B. *et al.* Water in a changing world. *Ecol. Appl.* **11**, 1027–1045 (2001).
4. Cramer, W. *et al.* Global response of terrestrial ecosystem structure and function to CO_2 and climate change: results from six dynamic global vegetation models. *Global Change Biol.* **7**, 357–373 (2001).
5. Ciais, P. *et al.* A three-dimensional synthesis study of δO^{18} in atmospheric CO_2 . I. Surface fluxes. *J. Geophys. Res. Atmos.* **102**, 5857–5872 (1997).
6. Edwards, D., Kerp, H. & Hass, H. Stomata in early land plants: an anatomical and ecophysiological approach. *J. Exp. Bot.* **49**, 255–278 (1998).
7. Beerling, D. J. & Woodward, F. I. Changes in land plant function over the Phanerozoic: reconstructions based on the fossil record. *Bot. J. Linn. Soc.* **124**, 137–153 (1997).
8. Woodward, F. I. Do plants really need stomata? *J. Exp. Bot.* **49**, 471–480 (1998).
9. Janis, C. M. Tertiary mammal evolution in the context of changing climates, vegetation, and tectonic events. *Annu. Rev. Ecol. Syst.* **24**, 467–500 (1993).
10. Mast, A. R. & Givnish, T. J. Historical biogeography and the origin of stomatal distributions in *Banksia* and *Dryandra* (Proteaceae) based on their cpDNA phylogeny. *Am. J. Bot.* **89**, 1311–1323 (2002).
11. Hill, R. S. (ed.) *History of the Australian Vegetation* (Cambridge Univ. Press, Cambridge, 1994).
12. Arnone, G. & De Micco, V. Seasonal dimorphism in the Mediterranean *Cistus incanus* L. subsp. *incanus*. *Ann. Bot.* **87**, 789–794 (2001).
13. Kellogg, E. A. Evolutionary history of the grasses. *Plant Physiol.* **125**, 1198–1205 (2001).
14. Palevitz, B. A. in *Stomatal Physiology* (eds Jarvis, P. G. & Mansfield, T. A.) 1–23 (Cambridge Univ. Press, Cambridge, 1981).
15. Raschke, K. in *Encyclopedia of Plant Physiology* Vol. 7 (eds Haupt, W. & Feinleib, M. E.) 383–441 (Springer, Berlin, 1979).
16. Grantz, D. A. & Assmann, S. M. Stomatal response to blue-light—water-use efficiency in sugarcane and soybean. *Plant Cell Environ.* **14**, 683–690 (1991).
17. Wever, L. A., Flanagan, L. B. & Carlson, P. J. Seasonal and interannual variation in evapotranspiration, energy balance and surface conductance in a northern temperate grassland. *Agric. For. Meteorol.* **112**, 31–49 (2002).
18. Retallack, G. J. Cenozoic expansion of grasslands and climatic cooling. *J. Geol.* **109**, 407–426 (2001).
19. Asanama, K., Sober, A. & Rahi, M. Leaf anatomical characteristics associated with shoot hydraulic conductance, stomatal conductance and stomatal sensitivity to changes of leaf water status in temperate deciduous trees. *Aust. J. Plant Physiol.* **28**, 765–774 (2001).
20. Meidner, H. & Mansfield, T. A. *Physiology of Stomata* (McGraw-Hill, London, 1968).
21. Allen, M. T. & Pearcy, R. W. Stomatal behavior and photosynthetic performance under dynamic light regimes in a seasonally dry tropical rain forest. *Oecologia* **122**, 470–478 (2000).
22. Woodward, F. I., Lake, J. A. & Quick, W. P. Stomatal development and CO_2 : ecological consequences. *New Phytol.* **153**, 477–484 (2002).
23. Royer, D. L. Stomatal density and stomatal index as indicators of paleoatmospheric CO_2 concentration. *Rev. Palaeobot. Palynol.* **114**, 1–28 (2001).
24. Gray, J. E. *et al.* The HIC signalling pathway links CO_2 perception to stomatal development. *Nature* **408**, 713–716 (2000).
25. Chen, X., Goodwin, M. S., Boroff, V. L., Liu, X. & Jenks, M. A. Cloning and characterization of the WAX2 gene of *Arabidopsis* involved in cuticle membrane and wax production. *Plant Cell* **15**, 1170–1185 (2003).
26. Lake, J. A., Woodward, F. I. & Quick, W. P. Long-distance CO_2 signalling in plants. *J. Exp. Bot.* **53**, 183–193 (2002).
27. Penninckx, I. A. M. A., Thomma, B. P. H. J., Buchala, A., Metraux, J. P. & Broekaert, W. F. Concomitant activation of jasmonate and ethylene response pathways is required for induction of a plant defense gene in *Arabidopsis*. *Plant Cell* **10**, 2103–2113 (1998).
28. Adam, L. *et al.* Comparison of *Erysiphe cichoracearum* and *E. cruciferarum* and a survey of 360 *Arabidopsis thaliana* accessions for resistance to these two powdery mildew pathogens. *Mol. Plant Microbe Interact.* **12**, 1031–1043 (1999).
29. Lake, J. A., Quick, W. P., Beerling, D. J. & Woodward, F. I. Signals from mature to new leaves. *Nature* **411**, 154–155 (2001).

30. Maldonado, A. M., Doerner, P., Dixon, R. A., Lamb, C. J. & Cameron, R. K. A putative lipid transfer protein involved in systemic resistance signalling in *Arabidopsis*. *Nature* **419**, 399–403 (2002).
31. Heichel, G. H. & Anagnostakis, S. L. Stomatal response to light of *Solanum pennellii*, *Lycopersicon esculentum*, and a graft-induced chimera. *Plant Physiol.* **62**, 387–390 (1978).
32. Schroeder, J. I., Allen, G. J., Hugouvieux, V., Kwak, J. M. & Waner, D. Guard cell signal transduction. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **52**, 627–658 (2001).
33. Hetherington, A. M. Guard cell signalling. *Cell* **107**, 711–714 (2001).
34. McAinsh, M. R. & Hetherington, A. M. Encoding specificity in Ca^{2+} signalling systems. *Trends Plant Sci.* **3**, 32–36 (1998).
35. Roelfsema, M. R. G. & Prins, H. B. A. Effect of abscisic acid on stomatal opening in isolated epidermal strips of *Arabidopsis thaliana*. *Physiol. Plant.* **95**, 373–378 (1995).
36. Webb, A. A. R. & Hetherington, A. M. Convergence of the abscisic acid, CO_2 , and extracellular calcium signal transduction pathways in stomatal guard cells. *Plant Physiol.* **114**, 1557–1560 (1997).
37. Leymarie, J., Lasceve, G. & Vavasseur, A. Interaction of stomatal responses to ABA and CO_2 on *Arabidopsis thaliana*. *Aust. J. Plant Physiol.* **25**, 785–791 (1998).
38. Leymarie, J., Vavasseur, A. & Lasceve, G. CO_2 sensing in stomata of *abi1-1* and *abi2-1* mutants of *Arabidopsis thaliana*. *Plant Physiol. Biochem.* **36**, 539–543 (1998).
39. Albert, R., Jeong, H. & Barabasi, A.-L. Error and attack tolerance of complex networks. *Nature* **406**, 378–382 (2000).
40. Strogatz, S. H. Exploring complex networks. *Nature* **410**, 268–276 (2001).
41. Barabasi, A.-L. *Linked the New Science of Networks* 280 (Perseus, Massachusetts, 2002).
42. Kinoshita, T. *et al.* phot1 and phot2 mediate blue light regulation of stomatal opening. *Nature* **414**, 656–660 (2001).
43. Hunt, L. *et al.* Phospholipase C is required for the control of stomatal aperture by ABA. *Plant J.* **34**, 47–55 (2003).
44. Webb, A. A. R., Larman, M. G., Montgomery, L. T., Taylor, J. E. & Hetherington, A. M. The role of calcium in ABA-induced gene expression and stomatal movements. *Plant J.* **26**, 351–362 (2001).
45. Frechilla, F., Talbott, L. D. & Zeiger, E. The CO_2 response of *Vicia* guard cells acclimates to growth environment. *J. Exp. Bot.* **53**, 545–550 (2002).
46. Medlyn, B. E. *et al.* Stomatal conductance of forest species after long-term exposure to elevated CO_2 concentration: a synthesis. *New Phytol.* **149**, 247–264 (2001).
47. Woodward, F. I. & Kelly, C. K. The influence of CO_2 concentration on stomatal density. *New Phytol.* **131**, 311–327 (1995).
48. Leymarie, J., Lasceve, G. & Vavasseur, A. Elevated CO_2 enhances stomatal responses to osmotic stress and abscisic acid in *Arabidopsis thaliana*. *Plant Cell Environ.* **22**, 301–308 (1999).
49. Webb, A. A. R. in *Biological Rhythms and Photoperiodism in Plants* (eds Lumsden, P. J. & Millar, A. J.) 69–79 (Bios, Oxford, 1998).
50. Talbott, L. D. & Zeiger, E. The role of sucrose in guard cell osmoregulation. *J. Exp. Bot.* **49**, 329–337 (1998).
51. Seidman, G. & Riggan, W. B. Stomatal movements: a yearly rhythm. *Nature* **217**, 684–685 (1968).
52. Thiel, G., MacRobbie, E. A. C. & Blatt, M. R. Membrane transport in stomatal guard cells—the importance of voltage control. *J. Membr. Biol.* **126**, 1–18 (1992).
53. Lohse, G. & Hedrich, R. Characterization of the plasma membrane H^+ + ATPase from *Vicia faba* guard cells modulation by extracellular factors and seasonal changes. *Planta* **188**, 206–214 (1992).
54. Willmer, C. & Sexton, R. Stomata and plasmodesmata. *Protoplasma* **100**, 113–124 (1979).
55. Leckie, C. P., McAinsh, M. R., Allen, G. J., Sanders, D. & Hetherington, A. M. Abscisic acid induced stomatal closure mediated by cyclic ADP ribose. *Proc. Natl Acad. Sci. USA* **95**, 15837–15842 (1998).
56. Weyers, J. D. B. & Lawson, T. Heterogeneity in stomatal characteristics. *Adv. Bot. Res.* **26**, 317–352 (1997).
57. Mott, K. A. & Buckley, T. N. Patchy stomatal conductance: emergent collective behaviour of stomata. *Trends Plant Sci.* **5**, 258–262 (2000).
58. Farquhar, G. D., Von Caemmerer, S. & Berry, J. A. A biochemical model of photosynthetic CO_2 assimilation in leaves of C3 species. *Planta* **149**, 78–90 (1980).
59. Jarvis, A. J., Mansfield, T. A. & Davies, W. J. Stomatal behaviour, photosynthesis and transpiration under rising CO_2 . *Plant Cell Environ.* **22**, 639–648 (1999).
60. Wulfschleger, S. D., Tschaplinski, T. J. & Norby, R. J. Plant water relations at elevated CO_2 —implications for water-limited environments. *Plant Cell Environ.* **25**, 319–331 (2002).
61. Woodward, F. I. Potential impacts of global elevated CO_2 concentrations on plants. *Curr. Opin. Plant Biol.* **5**, 207–211 (2002).
62. Lu, Z., Percy, R. G., Qualset, C. O. & Zeiger, E. Stomatal conductance predicts yields in irrigated Pima cotton and bread wheat grown in high temperatures. *J. Exp. Bot.* **49**, 453–460 (1998).
63. Dudley, S. A. Differing selection on plant physiological traits in response to environmental water availability: a test of adaptive hypotheses. *Evolution* **50**, 92–102 (1996).
64. Case, A. L. & Barrett, S. C. H. Ecological differentiation of combined and separate sexes of *Wurmbea dioica* (Colchicaceae) in sympatry. *Ecology* **82**, 2601–2616 (2001).
65. Avery, L. & Wasserman, S. Ordering gene function: the interpretation of epistasis in regulatory hierarchies. *Trends Genet.* **8**, 312–316 (1992).
66. Bredenkamp, G. J., Spada, F. & Kazmierczak, E. On the origin of northern and southern hemisphere grasslands. *Plant Ecol.* **163**, 209–229 (2002).
67. Sugden, A. M. Leaf anatomy in a Venezuelan montane forest. *Bot. J. Linn. Soc.* **90**, 231–241 (1985).
68. Tanner, E. V. J. & Kapos, V. Leaf structure of Jamaican upper montane rain-forest trees. *Biotropica* **14**, 16–24 (1982).
69. Knapp, A. K. Gas exchange dynamics on C3 and C4 grasses: consequences of differences in stomatal conductance. *Ecology* **74**, 113–123 (1993).
70. Li, W. L., Berlyn, G. P. & Ashton, P. M. S. Polyploids and their structural and physiological characteristics relative to water deficit in *Betula papyrifera* (Betulaceae). *Am. J. Bot.* **83**, 15–20 (1996).
71. Blanke, M. M., Höfer, M. & Pring, R. J. Stomata and structure of tetraploid apple leaves cultured *in vitro*. *Ann. Bot.* **73**, 651–654 (1994).
72. Mitton, J. B., Grant, M. C. & Yoshino, A. M. Variation in allozymes and stomatal size in pinyon (*Pinus edulis*, Pinaceae), associated with soil moisture. *Am. J. Bot.* **85**, 1262–1265 (1998).
73. Jones, H. G. Transpiration in barley lines with differing stomatal frequencies. *J. Exp. Bot.* **28**, 162–168 (1977).
74. Mishra, M. K. Stomatal characteristics at different ploidy levels in *Coffea* L. *Ann. Bot.* **80**, 689–692 (1997).
75. Knapp, A. K., Cocke, M., Hamerlynck, E. P. & Owensby, C. E. Effect of elevated CO_2 on stomatal density and distribution in a C4 grass and a C3 forb under field conditions. *Ann. Bot.* **74**, 595–599 (1994).
76. McConathy, R. K. Tulip-poplar leaf diffusion resistance calculated from stomatal dimensions and varying environmental parameters. *For. Sci.* **29**, 139–148 (1983).
77. Gindell, I. Stomata constellation in the leaves of cotton, maize and wheat plants as a function of soil moisture and environment. *Physiol. Plant.* **22**, 1143–1151 (1969).
78. Anderson, L. J., Maherali, H., Johnson, H. B., Polley, H. W. & Jackson, R. B. Gas exchange and photosynthetic acclimation over subambient to elevated CO_2 in a C3–C4 grassland. *Global Change Biol.* **7**, 693–707 (2001).
79. Cavender-Bares, J. & Bazzaz, F. A. Changes in drought response strategies with ontogeny in *Quercus rubra*: implications for scaling from seedlings to mature trees. *Oecologia* **124**, 8–18 (2000).
80. Ellsworth, D. S. CO_2 enrichment in a maturing pine forest: are CO_2 exchange and water status in the canopy affected? *Plant Cell Environ.* **22**, 461–472 (1999).
81. Escalona, M. H., Bota, J. M., Gulias, J. & Flexas, J. Regulation of photosynthesis of C-3 plants in response to progressive drought: stomatal conductance as a reference parameter. *Ann. Bot.* **89**, 895–905 (2002).
82. Fetene, M., Nauke, P., Lüttge, U. & Beck, E. Photosynthesis and photoinhibition in a tropical alpine giant rosette plant, *Lobelia rhynchopetalum*. *New Phytol.* **137**, 453–461 (1997).
83. Franco, A. C. & Lüttge, U. Midday depression in savanna trees: coordinated adjustments in photochemical efficiency, photorespiration, CO_2 assimilation and water use efficiency. *Oecologia* **131**, 356–365 (2002).
84. Garcia, R. L. *et al.* Photosynthesis and conductance of spring-wheat leaves: field response to continuous free-air atmospheric CO_2 enrichment. *Plant Cell Environ.* **21**, 659–669 (1998).
85. Giorio, P., Sorrentino, G. & d'Andria, R. Stomatal behaviour, leaf water status and photosynthetic response in field-grown olive trees under water deficit. *Environ. Exp. Bot.* **42**, 95–104 (1999).
86. Hamerlynck, E. P., Huxman, T. E., Charlet, T. N. & Smith, S. D. Effects of elevated CO_2 (FACE) on the functional ecology of the drought-deciduous Mojave Desert shrub. *Lycium andersonii*. *Environ. Exp. Bot.* **48**, 93–106 (2002).
87. Hirasawa, T. & Hsiao, T. C. Some characteristics of reduced leaf photosynthesis at midday in maize growing in the field. *Field Crops Res.* **62**, 53–62 (1999).
88. Huxman, T. E. & Smith, S. D. Photosynthesis in an invasive grass and native forb at elevated CO_2 during an El Niño year in the Mojave Desert. *Oecologia* **128**, 193–201 (2001).
89. Jiang, G. M. & Zhu, G. J. Different patterns of gas exchange and photochemical efficiency in three desert shrub species under two natural temperatures and irradiances in Mu Us Sandy Area of China. *Photosynthetica* **39**, 257–262 (2001).
90. Kaiser, H. & Kappen, L. *In situ* observation of stomatal movements and gas exchange of *Aegopodium podagraria* L. in the understory. *J. Exp. Bot.* **51**, 1741–1749 (2000).
91. Kazda, M., Salzer, J. & Reiter, I. Photosynthetic capacity in relation to nitrogen in the canopy of a *Quercus robur*, *Fraxinus angustifolia* and *Tilia cordata* flood plain forest. *Tree Physiol.* **20**, 1029–1037 (2000).
92. Knapp, A. K. Gas exchange dynamics in C3 and C4 grasses: consequences of differences in stomatal conductance. *Ecology* **74**, 113–123 (1993).
93. Lee, T. D., Tjoelker, M. G., Ellsworth, D. S. & Reich, P. B. Leaf gas exchange responses of 13 prairie grassland species to elevated CO_2 and increased nitrogen supply. *New Phytol.* **150**, 405–418 (2001).
94. Lodge, R. J., Dijkstra, P., Drake, B. G. & Morison, J. I. L. Stomatal acclimation to increased CO_2 concentration in a Florida scrub oak species *Quercus myrtifolia* Willd. *Plant Cell Environ.* **24**, 77–88 (2001).
95. McCarron, J. K. & Knapp, A. K. C3 woody plant expansion in a C4 grassland: are grasses and shrubs functionally distinct? *Am. J. Bot.* **88**, 1818–1823 (2001).
96. Medlyn, B. E., Loustau, D. & Delzon, S. Temperature response of parameters of a biochemically based model of photosynthesis. I. Seasonal changes in mature maritime pine (*Pinus pinaster* Ait.). *Plant Cell Environ.* **25**, 1155–1165 (2002).
97. Pääkkönen, E., Vahala, J., Pohjola, M., Holopainen, T. & Kärenlampi, L. Physiological, stomatal and ultrastructural ozone responses in birch (*Betula pendula* Roth.) are modified by water stress. *Plant Cell Environ.* **21**, 671–684 (1998).
98. Turnbull, M. H. *et al.* Photosynthetic characteristics in canopies of *Quercus rubra*, *Quercus prinus* and *Acer rubrum* differ in response to soil water availability. *Oecologia* **130**, 515–524 (2002).
99. Yoder, B. J., Ryan, M. G., Waring, R. H., Schoettle, A. W. & Kaufmann, M. R. Evidence of reduced photosynthetic rates in old trees. *For. Sci.* **40**, 513–527 (1994).
100. Yu, G. R., Zhuang, J. & Yu, Z. L. An attempt to establish a synthetic model of photosynthesis-transpiration based on stomatal behavior for maize and soybean plants grown in field. *J. Plant Physiol.* **158**, 861–874 (2001).

Acknowledgements A.M.H. wishes to acknowledge the support of the BBSRC to further stomatal research in his laboratory and the Fellows of St Catherine's College, Oxford, for the award of a Christensen Visiting Fellowship. F.I.W. is pleased to acknowledge support from NERC to further his stomatal research. We also wish to acknowledge G. Farquhar, J. Raven, M. Blatt, A. Webb, B. Davies, C. Price, N. Battey and J. Lake for providing input during the writing of this review, and L. Hunt, J. Gray and L. Mills for the stomatal images.

Correspondence and requests for materials should be addressed to A.M.H. (A.hetherington@lancaster.ac.uk).

Observing brownian motion in vibration-fluidized granular matter

G. D'Anna¹, P. Mayor¹, A. Barrat², V. Loreto³ & Franco Nori^{4,5}

¹Institut de Physique de la Matière Complexe, Faculté des Sciences de Base, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland

²Laboratoire de Physique Théorique, Unité Mixte de Recherche UMR 8627, Bâtiment 210, Université de Paris-Sud, 91405 Orsay Cedex, France

³Università degli Studi di Roma La Sapienza, Dipartimento di Fisica, and INFN, Center for Statistical Mechanics and Complexity, Piazzale A. Moro 5, 00185 Rome, Italy

⁴Frontier Research System, The Institute of Physical and Chemical Research (RIKEN), Wako-shi, Saitama 351-0198, Japan

⁵Center for Theoretical Physics, Department of Physics, CSCS, University of Michigan, Ann Arbor, Michigan 48109-1120, USA

Observation of the rotational brownian motion^{1,2} of a very fine wire immersed in a gas led to one of the most important ideas of equilibrium statistical mechanics. Namely, the many-particle problem of a large number of molecules colliding with the wire can be represented by just two macroscopic parameters: viscosity and temperature. Interest has arisen in the question of whether this idea (mathematically developed in the Langevin model and the fluctuation-dissipation theorem^{3,4}) can also be used to describe systems that are far from equilibrium. Here we report an experimental investigation of an archetypal non-equilibrium system, involving a sensitive torsion oscillator immersed in a granular system^{5,6} of millimetre-size grains that are fluidized by strong external vibrations. The vibro-fluidized granular medium is a driven environment, with continuous injection and dissipation of energy, and the immersed oscillator can be seen as analogous to an elastically bound brownian particle. By measuring the noise and the susceptibility, we show that the experiment can be treated (to a first approximation) with the equilibrium formalism. This gives experimental access to a granular viscosity and an effective temperature; however, these quantities are anisotropic and inhomogeneous. Surprisingly, the vibro-fluidized granular matter behaves as a 'thermal' bath satisfying a fluctuation-dissipation relation.

Before describing the experiment in the granular medium we briefly recall the Langevin formalism for the brownian motion of a torsion oscillator immersed in a usual liquid at temperature T . This is a canonical problem of statistical mechanics⁴. The Langevin equation of the oscillator is $I(d^2\theta/dt^2) + \alpha(d\theta/dt) + G\theta = C_{\text{ext}}(t) + R(t)$, where θ is the angular deflection, I is the moment of inertia of the oscillator, α the friction coefficient determining the viscous torque, G the fibre elastic torsion constant, $C_{\text{ext}}(t)$ the external torque, and $R(t)$ a randomly fluctuating torque, assumed to be a gaussian white noise of zero mean. This equation is valid on timescales large compared to the correlation time of the random torque.

The Langevin equation can be solved by Fourier transformation. Without the external torque ($C_{\text{ext}}(t) = 0$), using the Wiener-Khinchine theorem for stationary processes⁴, one obtains the noise power spectral density (PSD), that is, twice the Fourier transformation of the auto-correlation function $\langle \theta(t)\theta(t') \rangle$, which gives $S(\omega) = 2q/[I^2(\omega_0^2 - \omega^2)^2 + \alpha^2\omega^2]$ where $\omega_0 = \sqrt{G/I}$ is the natural pulsation of the oscillator and $2q$ is the PSD of the random torque $R(t)$. Moreover, one has the relation $q = 2\alpha k_B T$, which guarantees that the system is in thermodynamic equilibrium and that the equipartition of energy holds. On the other hand, solving the Langevin equation with the external torque $C_{\text{ext}}(t)$, and focusing on timescales large compared to the correlation time of the random torque, yields the complex susceptibility $\chi(\omega) = \theta(\omega)/C_{\text{ext}}(\omega) =$

$\chi'(\omega) - i\chi''(\omega)$ with $\chi(\omega) = 1/[I(\omega_0^2 - \omega^2) + i\alpha\omega]$. One finds that the ratio $S(\omega)\omega/(4\chi''(\omega)) = k_B T$ is proportional to the temperature of the liquid, and is independent of the oscillator characteristics, such as its mass or shape. This represents a formulation of the fluctuation-dissipation (FD) theorem⁴ (FDT), expressed in the frequency space. Summing up, measuring the susceptibility and the noise PSD of the brownian motion of the oscillator allows one to consistently determine the viscosity and the temperature of the liquid, seen as an ideal thermal bath at equilibrium.

The natural question that arises at this point is how much of this formalism survives if the experiment is performed in a strongly non-equilibrium system, notably in a vibro-fluidized granular medium^{5,6}. A granular system is an assembly of particles, such as sand grains or glass beads, interacting by contact forces and featuring a very large number of macroscopic degrees of freedom, corresponding in first approximation to the positions and velocities of all grains. However, such a system is not in 'equilibrium' in the thermodynamic sense, since the thermal energy $k_B T$ at room temperature is too small to induce any macroscopic grain fluctuations. Nevertheless, owing to the very large number of degrees of freedom, one expects that an analogy to the thermally induced brownian motion is possible when the system is externally driven and grain motion occurs by continuous injection of energy.

We have realized the experiment by immersing a torsion oscillator in a granular medium composed of glass beads (Fig. 1). The container, filled with the glass beads, is continuously shaken by a vertical vibrator, with a high-frequency filtered white noise, cut off below 300 Hz and above 900 Hz. Notice the important point that a filtered white noise is used in order to obtain a homogeneous and continuous agitation of the granular medium, discarding undesired effects such as pattern formation, rolls, and other instabilities in the granular motion⁵. The purpose is not to provide *ab initio* a random

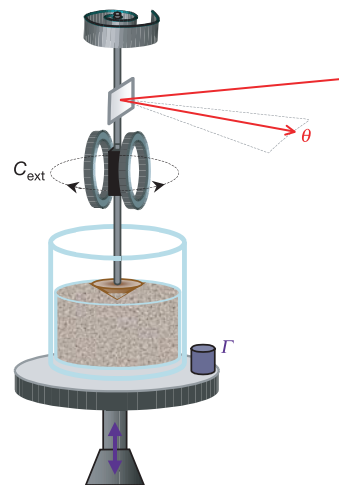


Figure 1 Sketch of the torsion oscillator immersed in a 'viscous liquid'. Here the 'viscous liquid' is a granular medium fluidized by strong external vibration (see text for details). The granular material consists of glass beads of diameter 1.1 ± 0.05 mm contained in a metallic bucket of 60 mm height and 94 mm diameter, filled to a height of 31 mm. The oscillator has a natural frequency of about 12 Hz, depending on the probe used. The main probe used is a cone of apex angle 120° , covered with a single layer of glass beads, glued on by epoxy adhesive. Probes with different shapes have also been used, in particular cylindrical probes with various diameters (see Fig. 3). The probe is immersed to depth L . The oscillator allows measurement of both the noise PSD and the complex susceptibility. An accelerometer permits measurement of the intensity of the external vibration, Γ . Notice that the oscillator does not move in the vertical or horizontal directions. For this reason, below $\Gamma \approx 1$ the oscillator acts as a localized perturbation point, and the study of the granular dynamics at low Γ requires a different approach^{14,15}.

torque with white noise spectrum to the oscillator. Indeed, we observe the motion of the oscillator in a low-frequency range (10–50 Hz) compared to the applied vibration high-frequency range. The torsion oscillator probe, with various moments of inertia and shapes, is immersed at a depth L from the granular surface. The oscillator is otherwise isolated from the container and the vibrator, and does not move in the vertical or horizontal directions. An accelerometer on the container is used to measure the acceleration spectrum, normalized to the acceleration of gravity, $A(\omega)$. We quantify the ‘intensity’ of the external shaking by Γ , defined as the square root of the band power of $A(\omega)$ in the range of about 1 Hz to 10 kHz, that is, $\Gamma^2 = \int A(f)df$ integrated in the above frequency range, with $\omega = 2\pi f$. In the case of a sinusoidal vibration of amplitude a_s and frequency $f_s = \omega_s/2\pi$, the normalized acceleration spectrum is $A(f) = [a_s(2\pi f_s)^2/g]^2 \delta(f - f_s)$ and one recovers the usual definition of the vibration intensity, that is, $\Gamma = a_s\omega_s^2/g$. For sinusoidal vibrations, $\Gamma = 1$ is the lift-on threshold above which a single grain starts to ‘fly’. In this work we employ vibration intensities up to $\Gamma \approx 15$.

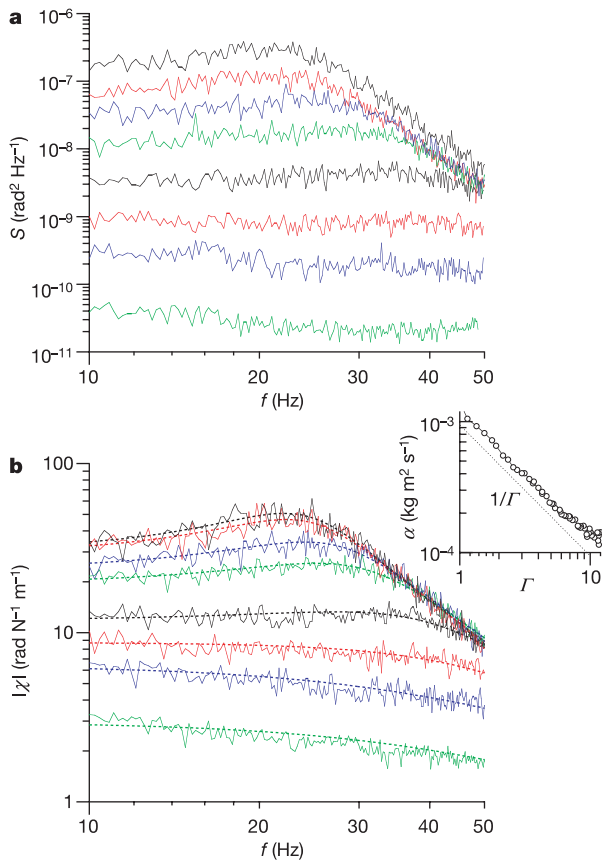


Figure 2 Noise and susceptibility for different vibration intensities. The probe is conical, immersed to about 11 mm from the surface. **a**, Noise PSD, $S(\omega)$, versus the frequency, f , for different vibration intensities Γ . The intensity values are, from top to bottom, 11.6, 10.0, 7.3, 5.4, 3.7, 2.2, 1.5 and 1.0. **b**, Modulus of the complex susceptibility $|\chi(\omega)|$ versus the frequency $f = \omega/2\pi$, for different vibration intensities Γ . The intensity values are as above. Each curve (continuous line) is fitted (dashed line) with the expression $|\chi(\omega)| = [I^2(\omega_0^2 - \omega^2)^2 + \alpha^2\omega^2]^{-1/2}$, with two free parameters, ω_0 and α . The moment of inertia of the oscillator is $I = 1.5 \times 10^{-6} \text{ kg m}^2$, and the applied torque amplitude is $C_0 = 3.2 \times 10^{-5} \text{ N m}$. The parameter ω_0 shows a small dependence on Γ : it features a small increase from 12 Hz to 18 Hz by reducing Γ . The inset shows the friction coefficient α , obtained from the fit to the curves $|\chi(\omega)|$ in the main panel, versus Γ . The dotted line is a power law $\alpha \propto 1/\Gamma$. The $1/\Gamma$ dependence is also systematically observed using susceptibility data obtained with probes of different shapes.

We now use the oscillator to measure (1) the noise PSD, and (2) the susceptibility of the vibro-fluidized granular medium. (1) In the absence of the external torque ($C_{\text{ext}}(t) = 0$) the immersed oscillator performs an irregular free angular motion, induced by the continuous interactions of the grains against the probe. The angular deflections θ are detected optically using a mirror fixed on the oscillator and, from the time-series of the angular deflections, $\theta(t)$, we obtain the noise PSD, $S(\omega)$, as shown in Fig. 2a for several values

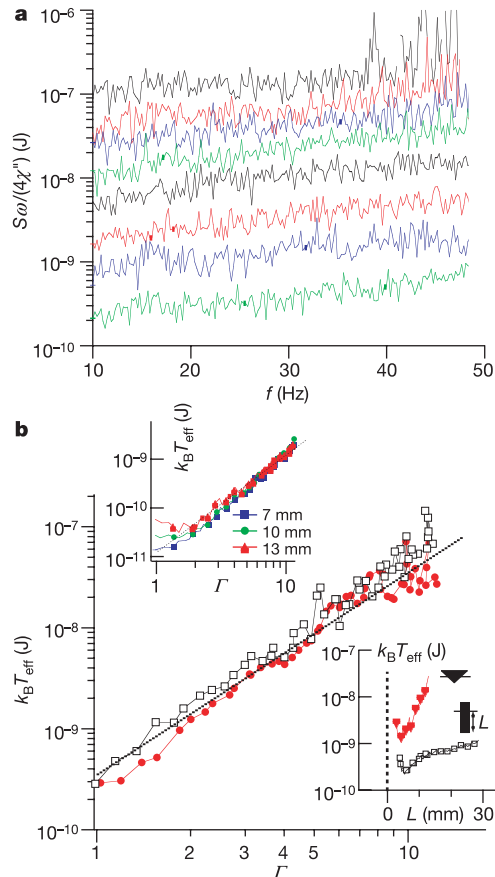


Figure 3 FD ratio and the effective temperature. **a**, The FD ratio $S(\omega)\omega/(4\chi''(\omega))$, versus $f = \omega/2\pi$ for different vibration intensities Γ , obtained from Fig. 2. The intensity values are, from top to bottom, 11.6, 10.0, 7.3, 5.4, 3.7, 2.2, 1.5 and 1.0. **b**, The FD ratio level (averaged from 10 Hz to 50 Hz), that is, the effective temperature T_{eff} , according to the expression $S(\omega)\omega/(4\chi''(\omega)) = k_B T_{\text{eff}}$, versus Γ , from data in **a** (black open symbols), and from similar data obtained using a conical probe with triple moment of inertia (red symbols) and the same L . A power law fitted to the data gives $T_{\text{eff}} \propto \Gamma^p$, with $p = 2.1$. The dashed line has equation $k_B T_{\text{eff}} = 3.5 \times 10^{-10} \Gamma^2$. The same FD ratio levels are measured by feeding the vibrator with a filtered white noise with a different central frequency (800 Hz, not shown). The FD ratio level also depends on the tribological properties of the granular material (especially rich for glass¹⁹). We have obtained reproducible data using aged (that is, exposed for a long time to the atmosphere) glass beads. Upper inset, the effective temperature T_{eff} versus Γ , obtained using cylindrical probes of different diameters and polished surfaces. (For clarity, only one symbol every four data values is marked). The immersion depth is about 21 mm. Notice the deviation from the $T_{\text{eff}} \propto \Gamma^2$ dependence at low Γ (see text for details). Lower inset, the effective temperature T_{eff} at $\Gamma = 9$, versus the immersion depth of the oscillator, L , for conical (red symbols) and cylindrical (black symbols) probes (see text for details) (For clarity, only one symbol every four data values is marked). Notice that the precise position of the surface is unknown for a vibrated granular medium, but it is approximately at $L = 0$ (vertical dashed line). The bottom of the container is at $L = 31$ mm. We note that, owing to boundary effects, in the layer close to the surface (about $0 < L < 5$ mm), the FD ratio is no longer ‘flat’, and we do not extract an effective temperature in this region.

of Γ . (2) While the granular medium is vibrated at a given intensity Γ , a sinusoidal torque $C_{\text{ext}}(t) = C_e \sin(\omega t)$ is applied to the oscillator using a permanent magnet fixed on the oscillator and two external coils. The complex susceptibility at the given ω is obtained from $\chi(\omega) = \theta(\omega)/C_{\text{ext}}(\omega)$, where $\theta(\omega)$ and $C_{\text{ext}}(\omega)$ are the Fourier transforms of $\theta(t)$ and $C_{\text{ext}}(t)$, respectively. The amplitude C_e of the external torque is small enough to be in the regime of linear response, that is, we measure the linear susceptibility. The complex susceptibility $\chi(\omega)$ versus ω is obtained by sweeping the frequency of the sinusoidal torque. The modulus of the complex susceptibility, $|\chi(\omega)|$, for different intensities of vibration Γ , is shown in Fig. 2b.

With both the susceptibility and the noise PSD data, we are now able to check whether the analogy to the thermally induced brownian motion makes sense. First of all, the modulus of the complex susceptibility, $|\chi(\omega)|$, is fitted remarkably well by the susceptibility expression for the damped oscillator, $|\chi(\omega)| = [I^2(\omega_0^2 - \omega^2)^2 + \alpha^2 \omega^2]^{-1/2}$, as shown in Fig. 2b. This means that the deterministic equation of motion, that is, $I(d^2\theta/dt^2) + \alpha(d\theta/dt) + G\theta = C_{\text{ext}}(t)$, describes well the response of the immersed oscillator, and allows us to define a granular friction coefficient α , or a granular viscosity $\mu \propto \alpha$. From the fitting parameters at different Γ we also deduce that $\alpha \propto 1/\Gamma$, as shown in the inset of Fig. 2b.

Second, we can obtain the FD ratio $S(\omega)\omega/(4\chi''(\omega))$ from our data. Figure 3a shows the FD ratio versus $f = \omega/2\pi$ at different Γ . The FD ratio is surprisingly 'flat', that is, approximately independent of the frequency in the observed low-frequency range (especially compared to what has been measured in other non-equilibrium thermal systems, such as in laponite⁷ and glycerol⁸.) This means that the high-frequency driven agitation of the granular medium acts on the oscillator as a source of random torque with white spectrum, at least in the 10–50 Hz range under investigation. Energy is thus injected at high-frequency, and spreads into a low-frequency white spectrum. A roughly flat FD ratio also provides support for the existence of an FD relation in off-equilibrium driven granular steady states. The FD ratio level can thus be used to define an effective temperature, T_{eff} . Figure 3b shows the averaged FD ratio levels, that is, $k_B T_{\text{eff}}$, versus Γ . Fitting to a power-law we obtain $k_B T_{\text{eff}} \propto \Gamma^p$ with p close to 2, thus giving $T_{\text{eff}} \propto \Gamma^2$.

Obviously, this is only a first-order approximation. Close inspection of Fig. 3a apparently shows a small frequency dependence. However, a useful torsion signal is only detected around the natural frequency of the oscillator, limiting our accessible frequency range. In future work, the study of the frequency dependence of the FD ratio over a large frequency range could give insights about possible energy 'cascade' in vibro-fluidized granular matter. Notice also that a deviation from the $T_{\text{eff}} \propto \Gamma^2$ dependence is seen close to the lift-on threshold, depending on the probe used (upper inset Fig. 3b). The main reason is that the immersed probe, which does not move in the vertical or horizontal directions, becomes a source of localized vibration at low Γ . In other words, at low Γ the granular medium is 'warmer' around the probe than in the bulk (the 'thermometer heats the sample'), and the Γ dependence of the effective temperature changes.

In the frame of this first-order approximation, our results suggest a simple picture. Owing to the complex dissipation processes between the grains, only a fixed fraction of the energy injected by the vibrator is effectively available as granular kinetic energy and is 'sensed' by the oscillator. In fact, we notice that the order of magnitude of the thermal energy, that is, $k_B T_{\text{eff}}$ as measured here, is consistent with realistic values of the mean kinetic energy per particle, as measured by grain-tracking methods (see, for example, refs 9 and 10). Thus, the effective temperature T_{eff} we measure seems to be related to the granular temperature, as usually defined in granular gases^{5,6}. Notice that the measured granular friction coefficient α decreases by increasing the effective temperature, since

$\alpha \propto 1/\Gamma$ and $T_{\text{eff}} \propto \Gamma^2$. This is the behaviour of liquids rather than gases. However, the oscillator sees an increasing effective temperature and a decreasing granular density as Γ is increased. A similar effect has been reported by Zik *et al.*¹¹ by observing the mobility of a sphere immersed in a vibro-fluidized granular medium.

We now study the dependence of T_{eff} on the oscillator properties, as well as on the characteristics of the granular medium, such as immersion depth and granular anisotropy. Figure 3b addresses all these issues. The main panel of Fig. 3b shows data obtained using a conical probe with different moments of inertia. The upper inset of Fig. 3b shows data obtained using cylindrical probes with different diameters. In all these circumstances it turns out that T_{eff} is insensitive to the change of these parameters. This is an indication that T_{eff} could be an intrinsic property of the granular medium.

The bottom inset of Fig. 3b shows data measured at different immersion depths, using conical and cylindrical probes. Two important aspects need to be underlined: First, in a layer close to the surface the measured FD ratio shows a strong frequency dependence, and it becomes 'flat' (as in Fig. 3a) only at a sufficient immersion depth. (The FD ratio becomes 'flat' approximately at the depth where a minimum appears in the bottom inset of Fig. 3b). Underneath the surface layer, we observe that T_{eff} increases with the depth. This is possibly related to a slight decrease of the granular density with the depth, as observed in simulations¹² and measurements¹³. The behaviour of the surface layer itself needs to be explored more carefully, as we cannot exclude a relation with density induced features (as observed in refs 10, 12 and 13). A second point is the difference between T_{eff} measured by conical and cylindrical probes. This is a peculiar granular effect and it is probably related to the intrinsic anisotropy of the system under gravity. Although the diameter of the horizontal section of the probe plays no role, the angle of inclination of the surface exposed to the granular agitation, relative to the vertical axis, seems to be a relevant parameter. As reported in ref. 10, the granular temperature is anisotropic, corresponding to horizontal and vertical impulse components. One thus expects in particular the transfer of the vertical impulse into torsion momentum to depend on the geometry and the surface state of the probe. Nevertheless, for a given kind of probe—for example, conical or cylindrical—the present experiment shows that one can satisfy the requirements for the Langevin approach to be applicable.

The most important finding obtained here is that it is possible to separate slow and fast degrees of freedom in vibro-fluidized granular matter and to model it, in the spirit of the Langevin approach, in terms of two intrinsic parameters: an effective friction coefficient and an effective temperature. In this perspective the oscillator indeed behaves as a brownian oscillator, and the vibro-fluidized granular matter can be seen as a macroscopic 'thermal bath'. We note that the analogy is only formal, since the granular medium is never in equilibrium in the thermodynamic sense. The experiment also shows that granular matter's mechanical anisotropy and inhomogeneity due to gravitation play an important role, inducing complications that are not present in usual gas or liquid media, which can be assumed to be isotropic and homogeneous. $\square$

Received 19 March; accepted 24 June 2003; doi:10.1038/nature01867.

- Gerlach, W. & Lehrer, E. Über die Messung der rotatorischen Brownschen Bewegung mit Hilfe einer Drehwaage. *Naturwissenschaften* **15**, 15 (1927).
- Uhlenbeck, G. E. & Goudsmit, S. A problem in Brownian motion. *Phys. Rev.* **34**, 145–151 (1929).
- Wang, M. C. & Uhlenbeck, G. E. On the theory of the Brownian motion II. *Rev. Mod. Phys.* **17**, 323–342 (1945).
- Kubo, R., Toda, M. & Hashitsume, N. *Statistical Physics II: Nonequilibrium Statistical Mechanics* (Springer, Berlin, 1991).
- Jaeger, H. M., Nagel, S. R. & Behringer, R. P. Granular solids, liquids and gases. *Rev. Mod. Phys.* **68**, 1259–1273 (1996).
- Mehta, A. & Halsey, T. C. (eds) Proceedings of the Workshop "Challenges in Granular Physics". *Adv. Complex Syst.* **4** (special issue), 287–534 (2001).
- Bellon, L., Ciliberto, S. & Laroche, C. Violation of the fluctuation-dissipation relation during the formation of a colloidal glass. *Europhys. Lett.* **53**, 511–517 (2001).

8. Grigera, T. S. & Israeloff, N. E. Observation of a fluctuation-dissipation-theorem violation in a structural glass. *Phys. Rev. Lett.* **83**, 5038–5041 (1999).
9. Dixon, P. K. & Durian, D. J. Speckle visibility spectroscopy and variable granular fluidization. *Phys. Rev. Lett.* **90**, 184302 (2003).
10. Yang, X., Huan, C., Candela, D., Mair, R. W. & Walsworth, R. L. Measurements of grain motion in dense, three-dimensional granular fluid. *Phys. Rev. Lett.* **88**, 044301 (2002).
11. Zik, O., Stavans, J. & Rabin, Y. Mobility of a sphere in vibrated granular media. *Europhys. Lett.* **17**, 315–319 (1992).
12. Barrat, A. & Loreto, V. Response properties in a model for granular matter. *J. Phys. A* **33**, 4401–4426 (2000).
13. Philippe, P. & Bideau, D. Compaction dynamics of a granular medium under vertical tapping. *Europhys. Lett.* **60**, 677–683 (2002).
14. D'Anna, G. & Gremaud, G. Vogel-Fulcher-Tammann type diffusive slowdown in weakly perturbed granular media. *Phys. Rev. Lett.* **87**, 254–302 (2001).
15. D'Anna, G., Mayor, P., Gremaud, G., Barrat, A. & Loreto, V. Extreme events-driven glassy behaviour in granular media. *Europhys. Lett.* **61**, 60–66 (2003).
16. Olivi-Tran, N., Fraysse, N., Girard, P., Ramonda, M. & Chatain, D. Modeling and simulations of the behavior of glass particles in a rotating drum in heptane and water vapor atmospheres. *Eur. Phys. J. B* **25**, 217–222 (2002).

Acknowledgements We thank Z. Racz and A. Vespignani for discussions and comments. A.B. and V.L. wish to thank the Ecole Polytechnique Fédérale de Lausanne for its hospitality during the realization of this work. F.B. thanks the National Science Foundation for support.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.D. (gianfranco.danna@epfl.ch).

Signature of optimal doping in Hall-effect measurements on a high-temperature superconductor

Fedor F. Balakirev¹, Jonathan B. Betts¹, Albert Migliori¹, S. Ono², Yoichi Ando² & Gregory S. Boebinger¹

¹National High Magnetic Field Laboratory, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

²Central Research Institute of Electric Power Industry, Komae, Tokyo 201-8511, Japan

High-temperature superconductivity is achieved by doping copper oxide insulators with charge carriers. The density of carriers in conducting materials can be determined from measurements of the Hall voltage—the voltage transverse to the flow of the electrical current that is proportional to an applied magnetic field. In common metals, this proportionality (the Hall coefficient) is robustly temperature independent. This is in marked contrast to the behaviour seen in high-temperature superconductors when in the ‘normal’ (resistive) state^{1–5}; the departure from expected behaviour is a key signature of the unconventional nature of the normal state, the origin of which remains a central controversy in condensed matter physics⁶. Here we report the evolution of the low-temperature Hall coefficient in the normal state as the carrier density is increased, from the onset of superconductivity and beyond (where superconductivity has been suppressed by a magnetic field). Surprisingly, the Hall coefficient does not vary monotonically with doping but rather exhibits a sharp change at the optimal doping level for superconductivity. This observation supports the idea that two competing ground states underlie the high-temperature superconducting phase.

In the 16 years since the discovery of high-temperature superconductors, much has been learned about the many unusual properties of these materials. Most prominently, the high-temperature superconducting state has been shown to be the first known

example of ‘*d*-wave’ superconductivity, in which the superconducting energy gap has the symmetry of a $d_{x^2-y^2}$ atomic orbital. The underlying mechanism that gives rise to this *d*-wave superconductivity is less well established. There is increasing evidence that the key role is played by an antiferromagnetic phase, in which neighbouring electron spins are anti-aligned, that is in competition with superconductivity. There remains, however, a significant lack of understanding about the ‘normal-state’ behaviour of the high-temperature superconductors, behaviour that is so consistently unusual that many believe the most promising route to a complete understanding of high-temperature superconductivity lies in an understanding of this abnormal ‘normal state’. Among the various abnormal normal-state properties, the Hall effect has been notoriously difficult to understand.

The properties of $\text{Bi}_2\text{Sr}_{2-x}\text{La}_x\text{CuO}_{6+\delta}$ (BSLCO) make it an ideal compound for a comprehensive study of the normal state at low temperatures. High-quality BSLCO single crystals can be produced over a wide range of carrier concentration by partially substituting Sr with La (ref. 7). Most importantly for these studies, a 60-T magnetic field, currently the practical limit for non-destructive generation of magnetic fields, is sufficient to completely suppress the superconductivity in BSLCO⁸. Single-crystal samples of BSLCO were prepared using the floating-zone technique⁷, and were exceptionally well characterized by measurements of resistivity, high-temperature Hall effect and thermopower among other techniques^{5,7–11}.

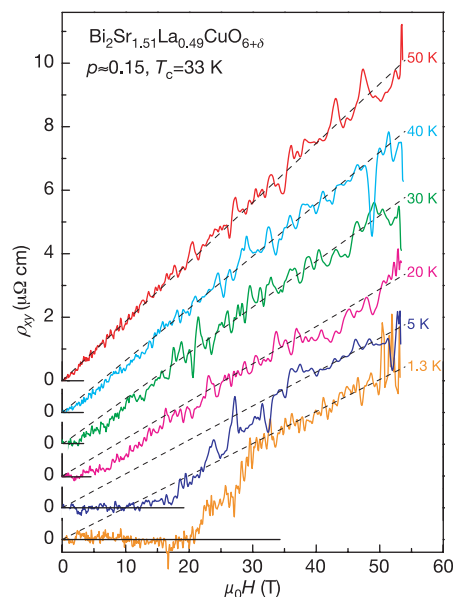


Figure 1 Hall resistivity versus magnetic field. Shown here is a typical Hall signal in a $\text{Bi}_2\text{Sr}_{0.51}\text{La}_{0.49}\text{CuO}_{6+\delta}$ superconducting sample for which $T_c = 33$ K and the hole doping $p \approx 0.15$. Measurements of the Hall resistivity were carried out during a magnetic field pulse of ~ 100 ms duration using a standard six-probe geometry for a series of ten samples. The data traces were recorded on a computer using a high-resolution low-noise synchronous lock-in technique developed at the National High Magnetic Field Laboratory. Data are shown for temperatures both above and below T_c , and the traces are offset vertically for clarity. The Hall resistivity above T_c shows the conventional linear magnetic-field dependence at all magnetic fields up to 55 T. At temperatures substantially below T_c , the Hall resistivity is zero at low magnetic field owing to the presence of the superconducting phase in the sample. As the magnetic field suppresses superconductivity, the Hall signal rapidly increases and then recovers its conventional linear-in-magnetic-field behaviour. The dashed lines represent best linear fit, $\rho_{xy}(H) = R_H H$, for the high-field regime in which superconductivity is suppressed by magnetic field, where R_H is the Hall coefficient, which, in a simple metal, is inversely proportional to the number of charge carriers in the material.

The sign of the Hall resistivity indicates that all the samples are 'hole-doped'—that is, charge transport is governed by electron vacancies, or 'holes', in the electron sea. The actual magnitude of the hole doping (p) for each sample is estimated by comparing BSLCO samples to other single-plane high-temperature superconductor compounds⁹. Optimum doping occurs at $p \approx 0.16$, where the superconducting transition temperature (T_c) is at a maximum. Figure 1 shows a typical field dependence of the Hall resistivity in a near-optimally-doped BSLCO sample with $p \approx 0.15$ and $T_c = 33$ K. We determine the low-temperature normal-state Hall coefficient (R_H) from a linear fit to the Hall resistivity data in

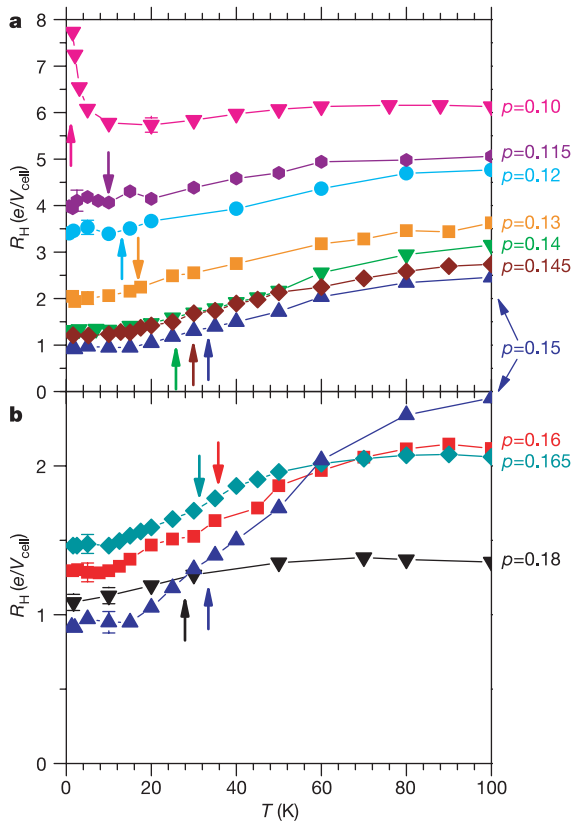


Figure 2 Temperature dependence of the Hall coefficient, R_H . R_H is determined from the slope of the high-field regime in $\text{Bi}_2\text{Sr}_{2-x}\text{La}_x\text{CuO}_{6+\delta}$ samples with different levels of hole doping (p). The La concentrations in the crystals have been determined by inductively coupled plasma analysis with 2% accuracy. The hole doping levels for samples with La concentration $x = 0.86, 0.80, 0.76, 0.68, 0.49, 0.45, 0.43, 0.34, 0.32$ and 0.24 are $p = 0.10, 0.115, 0.12, 0.13, 0.14, 0.145, 0.15, 0.16, 0.165$ and 0.18 , respectively, with an estimated 5% uncertainty. R_H has been normalized by the electric charge (e) and the unit cell volume (V_{cell}) in order to give the inverse Hall number per unit cell. In each sample, the intense magnetic fields have suppressed the superconducting transition temperature (T_c) from the zero-magnetic-field values denoted by the arrows in the figure. Above T_c these data overlay data previously taken in a low-field d.c. magnet on a similar set of samples¹¹. At low temperatures, R_H is nominally temperature-independent in all but the most underdoped sample ($p = 0.10$). In an ideal metal, the Hall number directly measures the density of charge carriers. If the same is true in BSLCO, the divergence of R_H in the $p = 0.10$ sample observed at low temperatures is consistent with a depletion of carriers, which is also evidenced by a diverging longitudinal resistivity consistent with carrier localization in this sample. Finally, we note that at high temperature (100 K), R_H evolves monotonically with p , whereas in the low-temperature limit R_H decreases with increasing p up to 0.15 , **a**, beyond which the behaviour becomes more complicated, **b**. For $p > 0.15$, **b**, there is a continuous reduction in the temperature dependence of R_H over the temperature range from 0 to 100 K as p increases, suggesting a recovery of conventional temperature-independent Hall behaviour in the overdoped limit.

the high-field regime, $\rho_{xy}(H) = R_H H$ (dashed lines in Fig. 1). We note that resistivity, $\rho_{xx}(H)$, measurements also show recovery of normal-state behaviour in pulsed magnetic fields⁸. This is consistent with a recent report that normal-state resistive transport can fully recover in high- T_c superconductors at magnetic fields smaller than the upper critical field H_{c2} (ref. 10). Because the linear-in-magnetic-field behaviour is observed at all temperatures and does not change up to the highest experimental fields, we conclude that the Hall transport experiments in the high-field linear regime measure the normal-state Hall resistivity at all temperatures. The temperature dependence of R_H for the ten BSLCO samples are plotted in Fig. 2. Note also that there is no noticeable change in the behaviour of $R_H(T)$ at T_c (arrows on Fig. 2), which provides further evidence that the $R_H(T)$ values in Fig. 2 correspond to the normal-state Hall coefficient in the absence of superconductivity. For all samples except the most underdoped sample, the value of R_H becomes relatively temperature independent below $T \approx 15$ K, signalling the recovery of conventional Hall behaviour once the inelastic scattering responsible for the anomalous temperature dependence 'freezes out' at low temperatures.

Figure 3 shows the surprising temperature and doping dependences of the Hall number per unit cell, $n_{\text{Hall}} \equiv V_{\text{cell}}/eR_H$. We focus initially on the evolution of n_{Hall} with doping in the low-temperature limit, the red curve in Fig. 3a, which increases rapidly from nearly zero at $p \approx 0.10$. This is the same value of p corresponding to the onset of the high-temperature superconducting state in BSLCO. n_{Hall} increases rapidly to approximately one hole per unit cell near

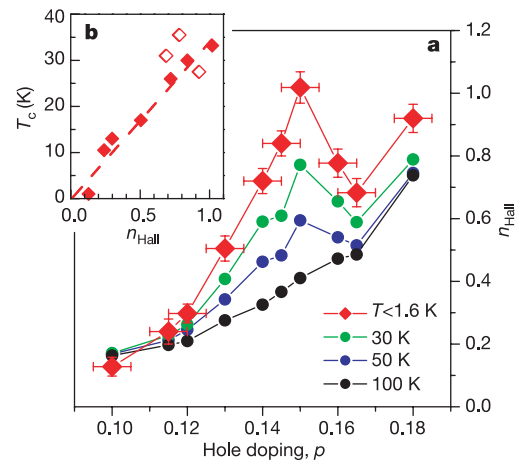


Figure 3 Variation of Hall number with doping and T_c . **a**, Hall number (n_{Hall}) variation with doping (p). n_{Hall} is determined from R_H values at various temperatures in ten BSLCO samples with doping levels ranging from underdoped ($p = 0.10$) to overdoped ($p = 0.18$). n_{Hall} values are normalized to give the number of holes per unit cell. The red curve represents the lowest-temperature behaviour of n_{Hall} . Note the rapid but continuous increase of low-temperature n_{Hall} from near zero at the onset of superconductivity at $p = 0.10$ (in fact n_{Hall} asymptotically approaches zero in the $p = 0.10$ sample upon cooling, Fig. 2a) to roughly 1 carrier per Cu at $p = 0.15$, even though the hole doping changed by only 0.05 holes per Cu. Note that the 100 K data (black line) show a monotonic evolution of the Hall number $n_{\text{Hall}}(p)$ with increasing doping, while upon cooling to ~ 50 K the first evidence of the anomalous cusp at $p = 0.15$ becomes apparent; this behaviour gives us confidence that the observed anomaly near optimum doping is intrinsic. This observation also highlights the necessity of high magnetic fields to reveal the intrinsic normal-state behaviour at low temperatures. The inset, **b**, shows the linear relationship between low temperature n_{Hall} and T_c for underdoped samples with $p \leq 0.15$ (solid diamonds), suggesting that the superfluid density corresponds to the normal carrier density in this doping range. Data for $p \geq 0.16$ samples are shown as open diamonds, and do not show any simple correlation between n_{Hall} and T_c .

optimal doping, resulting in roughly 7 carriers per each hole introduced with doping. At $p \approx 0.15$, there is a sharp change in the Hall number doping behaviour. This cusp suggests a sudden change in the Fermi surface of BSLCO at optimal doping.

Simple hole counting would suggest that the Hall number should depend linearly on p , equalling $1 + p$ over our entire range of hole doping. This is because the undoped ($p = 0$) copper oxide compound has one hole per unit cell in the highest occupied energy level. Although a partially filled energy level might normally be expected to result in metallic behaviour, the electron–electron repulsion fixes one electron on each copper atom. Electron motion is inhibited by the huge energy cost of double-electron occupancy on a single copper atom, resulting in the so-called Mott insulating state. Doping a Mott insulator eventually breaks down the localizing effects of the electron–electron interaction, although it is not understood how rapidly the Hall number should increase from zero to a finite number of order $1 + p$.

The maximum value of n_{Hall} that we observe corresponds to almost exactly one hole per unit cell, which implies a big Fermi surface near optimal doping, close to half a Brillouin zone in size and comparable to what has been seen in angle-resolved photoemission spectroscopy (ARPES) studies in similar high-temperature superconductors^{12,13} and in recent studies of modulations of differential tunnelling conductance¹⁴. In the underdoped regime, ARPES data show a disappearance of portions of the Fermi surface¹⁵, a phenomenon which is not yet understood and, thus, difficult to meaningfully compare to our data in the underdoped regime, although n_{Hall} might be a measure of the remaining parts of the Fermi surface at various levels of doping. In fact, scanning tunnelling microscopy experiments show clear evidence of spatially separated conducting and insulating regions in samples throughout the underdoped regime¹⁶. Because our Hall experiments might only be measuring those carriers which occupy the metallic regions that are connected percolatively throughout the sample, a rapid but continuous transition from zero to roughly one carrier per unit cell might be expected for n_{Hall} .

The rapid increase in Hall number through the underdoped regime ($p \leq 0.15$) shows a remarkably linear correlation with the doping dependence of T_c (Fig. 3b). Considering the well-known linear relation between the superfluid density and T_c , which has been established to be generic to the underdoped copper oxides^{17,18}, it is reasonable to conclude from our results that the superfluid density in the superconducting state corresponds to the carrier density in the underlying normal state throughout the underdoped regime. This correspondence becomes apparent only after high magnetic fields have revealed the zero-temperature limiting values of n_{Hall} .

It is illuminating that the normal-state Hall coefficient at sufficiently low temperature becomes free from complicated inelastic scattering and directly reflects the underlying electronic structure. The pronounced cusp in the zero-temperature limiting value of the Hall number at optimal doping must be associated with an abrupt change in the Fermi surface, and thus suggests a quantum phase transition (QPT) between two different metallic phases in the normal state of the high-temperature superconductors—as has been proposed on the basis of experiments on superconducting samples¹⁹ and which is the subject of active debate among theorists^{20–25}. A recent example of sharp changes in Hall number behaviour at a QPT in a common metal is provided by $\text{Cr}_{1-x}\text{V}_x$ alloy: as V doping increases, this system undergoes a zero-temperature quantum transition from an antiferromagnetic to a non-magnetic phase. The transition manifests itself as a non-analytical discontinuity in the low-temperature Hall number as a function of V doping²⁶, a behaviour similar to the one we observe in BSLCO.

A remarkable example of superconductivity induced by fluctuations of magnetism near a QPT boundary has been demonstrated in heavy-fermion compounds²⁷; when pressure is applied to these

systems, a superconducting phase is observed to be centred on a QPT. Another example is β -vanadium bronze, where superconductivity appears around the point of collapse of a charge-ordered phase²⁸. In light of these other systems, it is striking that we also find evidence of a QPT exactly at optimal conditions for superconductivity. We also note that our evidence for a QPT occurs where the linear temperature dependence of the resistivity is most robust²². In the double-layer ruthenate, linear resistivity is found to be an experimental hallmark of a QPT even in the absence of superconductivity²⁹. From these similar behaviours across a wide range of different materials, one can argue that the occurrence of high-temperature superconductivity is fundamentally related to quantum fluctuations associated with a zero-temperature phase transition. One possibility why the phase diagram for the high-temperature superconductors is still so much in debate arises from the extreme robustness of the superconductivity: once the temperature is increased enough to destroy the superconductivity, the evidence of a sharp phase transition is obscured by thermal fluctuations. It takes an extremely high magnetic field to suppress the superconductivity and reveal a sharp change in the normal state in the zero-temperature limit. $\square$

Received 20 September 2002; accepted 7 July 2003; doi:10.1038/nature01890.

1. Clayhold, J., Ong, N. P., Wang, Z. Z., Tarascon, J. M. & Barbour, P. Hall-effect anomaly in the high- T_c copper-based perovskites. *Phys. Rev. B* **39**, 7324–7327 (1989).
2. Anderson, P. W. Hall-effect in the 2-dimensional Luttinger liquid. *Phys. Rev. Lett.* **67**, 2092–2094 (1991).
3. Chien, T. R., Wang, Z. Z. & Ong, N. P. Effect of Zn impurities on the normal-state Hall angle in single-crystal $\text{YBa}_2\text{Cu}_3\text{--}x\text{Zn}_x\text{O}_{7-\delta}$. *Phys. Rev. Lett.* **67**, 2088–2091 (1991).
4. Hwang, H. Y. *et al.* Scaling of the temperature dependent Hall effect in $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$. *Phys. Rev. Lett.* **72**, 2636–2639 (1994).
5. Ando, Y. & Murayama, T. Nonuniversal power law of the Hall scattering rate in a single-layer cuprate $\text{Bi}_2\text{Sr}_{2-x}\text{La}_x\text{CuO}_6$. *Phys. Rev. B* **60**, 6991–6994 (1999).
6. Orenstein, J. & Millis, A. J. Advances in the physics of high-temperature superconductivity. *Science* **288**, 468–474 (2000).
7. Ono, S. & Ando, Y. Evolution of the resistivity anisotropy in $\text{Bi}_2\text{Sr}_{2-x}\text{La}_x\text{CuO}_{6+\delta}$ single crystals for a wide range of hole doping. *Phys. Rev. B* **67**, 104512 (2003).
8. Ono, S. *et al.* Metal-to-insulator crossover in the low-temperature normal state of $\text{Bi}_2\text{Sr}_{2-x}\text{La}_x\text{CuO}_{6+\delta}$. *Phys. Rev. Lett.* **85**, 638–641 (2000).
9. Ando, Y. *et al.* Carrier concentrations in $\text{Bi}_2\text{Sr}_{2-x}\text{La}_x\text{CuO}_{6+\delta}$ single crystals and their relation to the Hall coefficient and thermopower. *Phys. Rev. B* **61**, 14956–14959 (2000).
10. Wang, Y. *et al.* Dependence of upper critical field and pairing strength on doping in cuprates. *Science* **299**, 86–89 (2003).
11. Ando, Y., Murayama, T. & Ono, S. Systematic evolution of the magnetotransport properties of $\text{Bi}_2\text{Sr}_{2-x}\text{La}_x\text{CuO}_6$ in a wide doping range. *Physica C* **341**, 1913–1914 (2000).
12. Sato, T. *et al.* High-resolution angle-resolved photoemission study of Pb-substituted Bi2201 . *J. Phys. Chem. Solids* **62**, 157–161 (2001).
13. Sato, T. *et al.* Evidence for a hole-like Fermi surface of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$ from temperature-dependent angle-resolved photoemission spectroscopy. *Phys. Rev. B* **64**, 054502 (2001).
14. Hoffman, J. E. *et al.* Imaging quasiparticle interference in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Science* **297**, 1148–1151 (2002).
15. Norman, M. R. *et al.* Destruction of the Fermi surface in underdoped high- T_c superconductors. *Nature* **392**, 157–160 (1998).
16. Pan, S. H. *et al.* Microscopic electronic inhomogeneity in the high- T_c superconductor $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Nature* **413**, 282–285 (2001).
17. Uemura, Y. J. *et al.* Universal correlations between T_c and n_s/m^* (carrier density over effective mass) in high- T_c cuprate superconductors. *Phys. Rev. Lett.* **62**, 2317–2320 (1989).
18. Uemura, Y. J. *et al.* Basic similarities among cuprate, bismuthate, organic, chevron-phase, and heavy-fermion superconductors shown by penetration-depth measurements. *Phys. Rev. Lett.* **66**, 2665–2668 (1991).
19. Tallon, J. L. *et al.* Critical doping in overdoped high- T_c superconductors: a quantum critical point? *Phys. Status Solidi B* **215**, 531–540 (1999).
20. Chakravarty, S., Nayak, C., Tewari, S. & Yang, X. Sharp signature of $d_{x^2-y^2}$ quantum critical point in the Hall coefficient of cuprate superconductors. *Phys. Rev. Lett.* **89**, 277003 (2002).
21. Sachdev, S. Quantum criticality: Competing ground states in low dimensions. *Science* **288**, 475–480 (2000).
22. Varma, C. M., Littlewood, P. B., Schmitt-Rink, S., Abrahams, E. & Ruckenstein, A. E. Phenomenology of the normal state of Cu-O high-temperature superconductors. *Phys. Rev. Lett.* **63**, 1996–1999 (1989).
23. Perali, A., Castellani, C., Di Castro, C. & Grilli, M. *d*-wave superconductivity near charge instabilities. *Phys. Rev. B* **54**, 16216–16225 (1996).
24. Kivelson, S. A., Fradkin, E. & Emery, V. J. Electronic liquid-crystal phases of a doped Mott insulator. *Nature* **393**, 550–553 (1998).
25. Lee, P. A. & Wen, X.-G. Vortex structure in underdoped cuprates. *Phys. Rev. B* **63**, 224517 (2001).
26. Yeh, A. *et al.* Quantum phase transition in a common metal. *Nature* **419**, 459–462 (2002).
27. Mathur, N. D. *et al.* Magnetically mediated superconductivity in heavy fermion compounds. *Nature* **394**, 39–43 (1998).
28. Yamauchi, T., Ueda, Y. & Mori, N. Pressure-induced superconductivity in $\beta\text{-Na}_{0.33}\text{VO}_5$ beyond charge ordering. *Phys. Rev. Lett.* **89**, 057002 (2002).

29. Grigera, S. A. *et al.* Magnetic field-tuned quantum criticality in the metallic ruthenate $\text{Sr}_3\text{Ru}_2\text{O}_7$. *Science* **294**, 329–332 (2001).

Acknowledgements The work at the National High Magnetic Field Laboratory was supported by the National Science Foundation and the DOE Office of Science. We thank S. Chakravarty, S. A. Kivelson, P. A. Lee, R. Ramazashvili, C. M. Varma, I. Vekhter and F.-C. Zhang for discussions.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.F.B. (fedor@lanl.gov).

Epoxidation of polybutadiene by a topologically linked catalyst

Pall Thordarson, Edward J. A. Bijsterveld, Alan E. Rowan & Roeland J. M. Nolte

Department of Organic Chemistry, NSRIM, University of Nijmegen, Toernooiveld 1, 6525 ED, Nijmegen, The Netherlands

Nature has evolved complex enzyme architectures that facilitate the synthesis and manipulation of the biopolymers DNA and RNA, including enzymes capable of attaching to the biopolymer substrate and performing several rounds of catalysis before dissociating^{1–5}. Many of these ‘processive’ enzymes have a toroidal shape and completely enclose the biopolymer while moving along its chain, as exemplified by the DNA enzymes T4 DNA polymerase holoenzyme⁶ and λ -exonuclease⁷. The overall architecture of these systems resembles that of rotaxanes, in which a long molecule or polymer is threaded through a macrocycle. Here we describe a rotaxane that mimics the ability of processive enzymes to catalyse multiple rounds of reaction while the polymer substrate stays bound. The catalyst consists of a substrate binding cavity incorporating a manganese(III) porphyrin complex that oxidizes alkenes within the toroid cavity, provided a ligand has been attached to the outer face of the toroid to both activate the porphyrin complex and shield it from being able to oxidize alkenes outside the cavity. We find that when threaded onto a polybutadiene polymer strand, this catalyst epoxidizes the double bonds of the polymer, thereby acting as a simple analogue of the enzyme systems.

Rotaxanes and the topologically related catenanes are of interest since these mechanically interlocked assemblies represent a unique form of bonding. They have been shown to function as “molecular motors”^{8–11}, “molecular shuttles”¹², “molecular muscles”¹³, and even as active components in electronic devices¹⁴. Polyrotaxanes can be prepared by statistically threading toroids, namely cyclodextrins and crown ethers, onto polymers such as polyethyleneglycol¹⁵ and polyurethane¹⁶. Until now, no examples of catalytically active rotaxanes have been reported. With the natural processive enzymes in mind, we have designed a rotaxane catalytic system in which a cavity containing porphyrin (**Mn1**) is threaded onto a polybutadiene polymer (PD) and can move along it while catalysing the conversion of the double bonds in the thread into the corresponding epoxide functions (Fig. 1a).

The toroidal catalyst used is based on a glycoluril clip molecule^{17,18} that is capped with a manganese(III) porphyrin complex (**Mn1**, Figs 1 and 2), resulting in a macrocyclic compound with a central cavity in which substrates can bind. Manganese(III) porphyrins have been investigated extensively as oxidative catalysts, most prominently in the epoxidation of alkenes¹⁹. In the presence of an oxygen donor, for example, iodosylbenzene (PhIO) or sodium

hypochlorite (NaOCl) and an activating axial ligand, a $\text{Mn(V)} = \text{O}$ species is formed, which transfers its oxygen to the substrate (alkene) to form the product¹⁹.

The porphyrin **Mn1** has previously been shown to be an excellent catalyst for the conversion of alkenes to epoxides²⁰. In the presence of a small ligand such as pyridine (**py**), which binds in the cavity, catalysis occurs on the ‘outside’ of this catalyst (approach A in Fig. 1b). When a bulky axial ligand such as *tert*-butylpyridine (**tbpy**), which can only bind to the ‘outside’ of **Mn1**, is used, the catalysis occurs predominantly inside the cavity (approach B in Fig. 1b). To investigate whether **Mn1** could mimic the action of

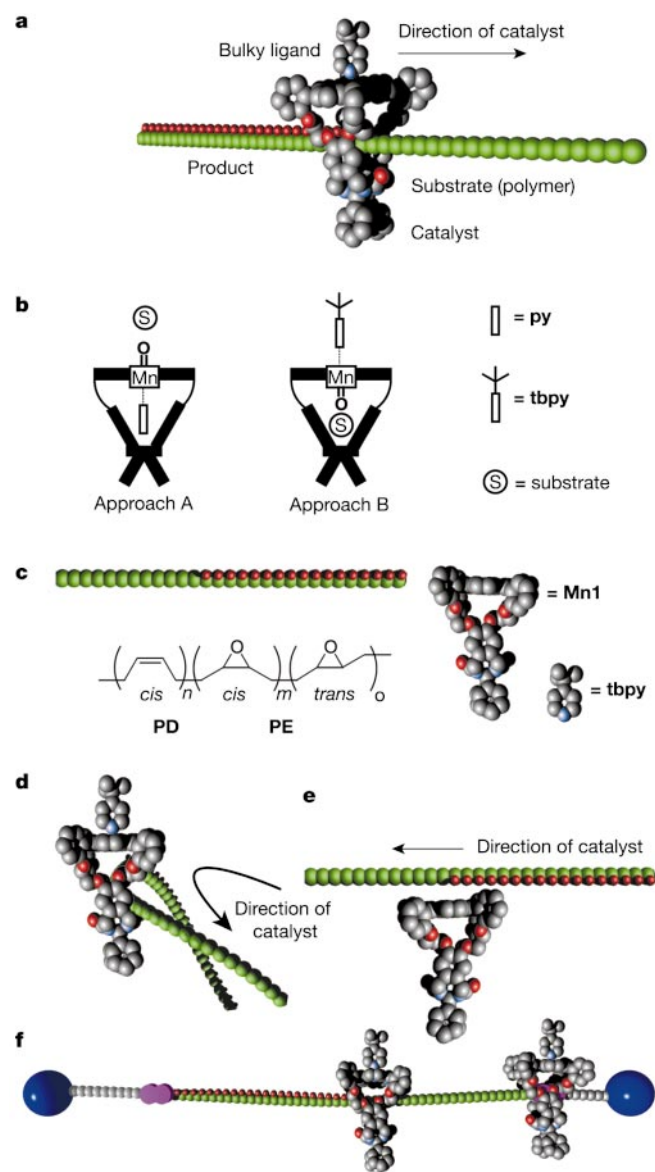


Figure 1 The various catalytic architectures discussed in this work. **a**, Schematic representation of the toroidal catalyst **Mn1** encircling a polybutadiene substrate and converting it into the product. **b**, Two ways in which catalyst **Mn1** can act as an epoxidation catalyst in combination with ligands **py** and **tbpy**. **c**, Structural representation of polybutadiene **PD**, the product **PE**, the **Mn1** catalyst, and the ligand **tbpy**. **d**, Schematic representation of the ‘loop’ mechanism where the substrate folds into the cavity of **Mn1**. **e**, Schematic representation of the ‘outside’ mechanism where the substrate reacts on the ‘outside’ (top) of **Mn1**. **f**, Schematic representation of the catalytically active rotaxane **Mn3**.

processive enzymes and carry out catalysis on a polymer in a rotaxane-like architecture, that is, in which the polymer substrate is threaded through the macrocycle, this complex was tested as a catalyst for the conversion of polybutadiene (PD) into polybutadieneepoxide (PE)²¹. In the presence of **tbpy** as ligand, which forces catalysis to occur in the cavity, and PhIO as an oxidant, the relative rates of conversion of PD into PE were compared to **Mn2** (Table 1, entries 1–5). This latter catalyst is electronically related to **Mn1** but does not possess a binding cavity. ¹H NMR analysis of the products (only epoxides were formed²²) revealed that **Mn1** is a slower catalyst (by a factor of 2.2, compare entries 1 and 3) than **Mn2** for the conversion of PD to PE. This reduced activity is not unexpected since the reaction of PD is forced to take place inside the cavity of **Mn1** (Fig. 1a, b). When the cavity of **Mn1** was intentionally blocked by the addition of 1 equiv. of viologen **V** ($K_{\text{ass}} > 10^5 \text{ M}^{-1}$)¹⁸, which acts as an inhibitor (Supplementary Information), the reaction rate dropped significantly (by ~40%, compare entries 1 and 2). This is in contrast to control experiments with **Mn2**, where it was shown that **V** actually activates this simple catalyst (compare entries 3 and 4)²³.

When the concentration of all the reagents was lowered, the rate of epoxidation using **Mn1** dropped accordingly (compare entries 1 and 5 in Table 1), see below. Moreover, while toroidal catalyst **Mn1** produced 80% *trans*- and 20% *cis*-epoxide polymer from PD (>98% *cis*-butadiene), **Mn2** gave predominantly the *cis*-product (78% *cis*, 22% *trans*, see Table 2). This large difference in stereoselectivity strongly suggests that in the case of **Mn1**, catalysis occurs in the sterically demanding cavity, as shown previously in our work with simpler substrates such as stilbenes²⁰.

Taken together, these observations suggest that **Mn1** forms a pseudo-rotaxane complex with the PD polymer and then acts as a topologically linked catalyst (Fig. 1a). Two other possible mechanisms could also (partly) explain the experimental results. One

involves folding of the polymer substrate and bending inside the cavity of **Mn1** (Fig. 1d), the other, reaction of PD on the 'outside' of **Mn1** (Fig. 1e). Simple molecular modelling suggests that the first of these two possibilities is unlikely owing to the large steric bulk of the polymer, while the second possibility should only play a minor role in the presence of an excess of **tbpy** (Supplementary Information).

To investigate the scope for PD to PE conversion inside the cavity of **Mn1** more directly, we designed and synthesized a polymer-porphyrin rotaxane **Mn3** as a model compound. In this molecule, the rotaxane architecture with the catalyst enclosing the polybutadiene substrate is enforced. The first alternative mechanism, the 'loop' mechanism, is thus sterically impossible; moreover, catalysis on the 'outside' of the catalyst is topologically inhibited.

The synthesis of **Mn3** is based on our earlier work in which the strong affinity of **Mn1** for **V** is used to form a pseudo-rotaxane structure that is subsequently capped with a suitable stopper to form a rotaxane¹⁷. Using this approach, bis-bipyridyl-appended polybutadiene (**bipy₂PD**) was treated with 2 equiv. of porphyrin **Mn1** to yield pseudo-rotaxane **Mn1:bipy₂PD:Mn1**, which was subsequently capped with a 3,5-di-*tert*-butylphenyl derivative to obtain the target polymer-porphyrin-[3]-rotaxane **Mn3** in approximately 43% yield (Supplementary Information).

Catalytic studies using the [3]-rotaxane **Mn3** both in the presence and in absence of the **tbpy** ligand were carried out as described above and the results compared with those obtained using the porphyrin clips **Mn1** and **Mn2**. It should be noted that in the case of rotaxane **Mn3**, the substrate (polybutadiene) is now part of the molecule, whereas in the other cases the related polybutadiene **bipy₂PD** was used as the substrate. The substrate/catalyst ratio in the pseudo-rotaxane systems is fixed by the ratio of the two components in the **Mn3** rotaxane, which is ~40. We used substrate concentrations below 10 mM, thus suppressing intermolecular side reactions but also obtaining lower turnover numbers (Table 1,

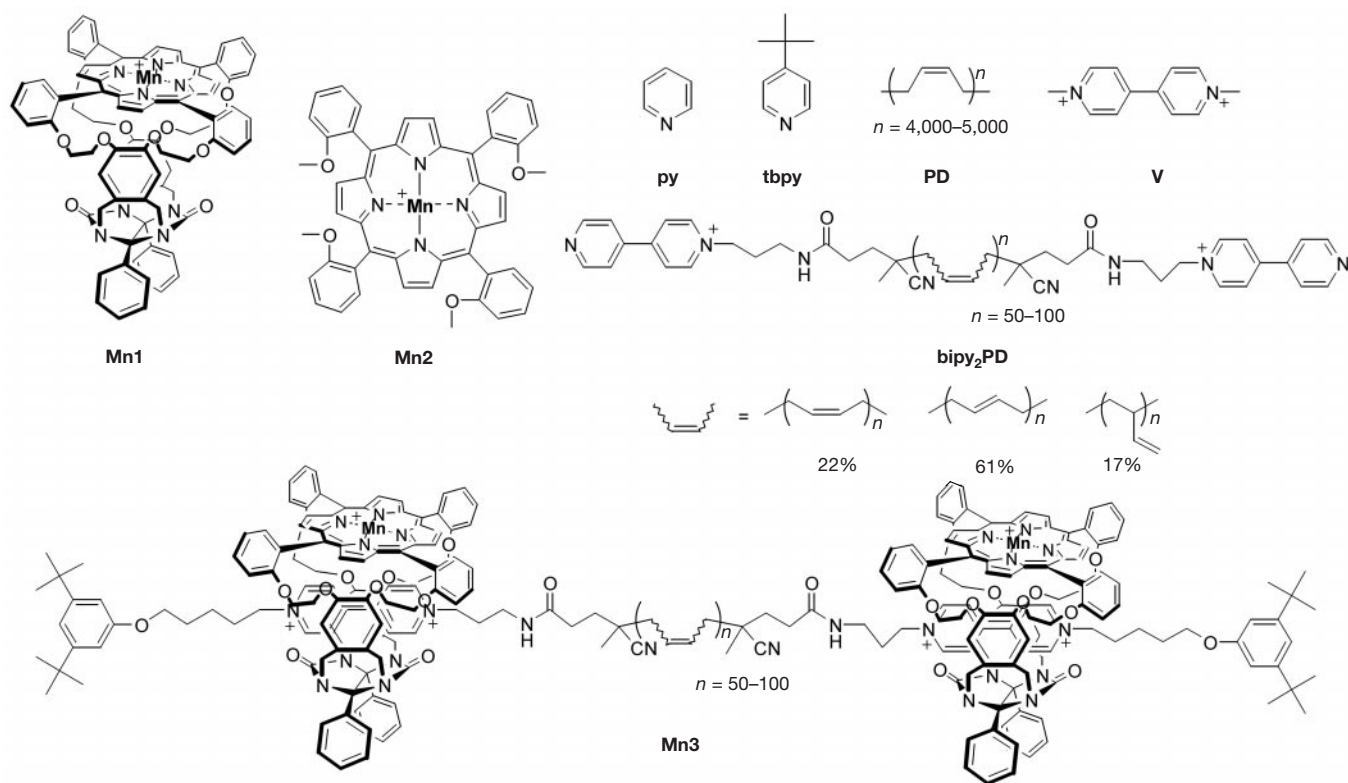


Figure 2 Structural formulae of the catalysts, substrates and ligands used in this work.

Table 1 Catalytic activities of manganese(III) porphyrins

Entry	Catalyst	Substrate	[C = C] (mM)	[C = C]/[catalyst]	Turnover number (h)	
					tbpy	No ligand
1	Mn1	PD	250	250	140	—
2	Mn1:V*	PD	250	250	85	—
3	Mn2	PD	250	250	310	—
4	Mn2:V*	PD	250	250	375	—
5	Mn1	PD	8	42	15	21
6	Mn1	bipy₂PD	8	43	7	20
7	Mn2	bipy₂PD	8	42	21	10
8	Mn3	Mn3†	6	37	12	5
9	Zn1‡	bipy₂PD	8	45	0	—

[C = C] = concentration of double bonds in **PD** (Acros, $M_w \sim 300,000$, 98% *cis*), **bipy₂PD** (Aldrich, $M_w \sim 4,500$; 22% 1,4-*cis*-, 61% 1,4-*trans*-, and 17% 1,2-butadiene) and **Mn3** ($M_w \sim 9,000$, based on **bipy₂PD**). See Fig. 2 for structures and abbreviations and Methods for experimental details. Turnover number = ([substrate converted]/[catalyst]). For **Mn3**, [catalyst] = concentration of Mn(III) porphyrin ($2 \times [\text{Mn3}]$).

*1 equiv. of **V** added.

†In **Mn3**, the catalyst and substrate are topologically linked.

‡**Zn1** is the zinc(II) analogue of **Mn1**¹⁸ and was used as a control to prove that PhIO does not carry out oxidation reactions in the absence of manganese(III) porphyrin.

entries 5–9). The activity of our system is, however, still comparable to that measured for naturally occurring oxidative systems such as the enzyme Cytochrome P450 (0.1–60 turnovers per catalyst per hour)^{24,25}.

In the presence of the **tbpy** ligand, the polymeric porphyrin-rotaxane **Mn3** is a more effective catalyst ($\sim 2 \times$ higher turnovers per hour) than the pseudo-rotaxane system **Mn1-bipy₂PD** (Table 1, entries 6 and 8). We attribute the increased activity to the enforced rotaxane topology of **Mn3** (Fig. 1f), which ensures that the initial binding of the catalyst to the substrate is already achieved. In line with the inhibition experiments discussed above, the turnover frequency for the conversion of **bipy₂PD** by **Mn1** is 50% lower than that of **PD** by **Mn1** (Table 1, entries 5 and 6), indicating that the bipyridine-endgroups of the **bipy₂PD** substrate bind strongly in the cavity and inhibit catalysis (compare to Table 1, entries 1 and 2). The catalytic turnover of **Mn1** is lower in the presence of **tbpy** than in the absence of **tbpy** ($\sim 65\%$ lower, entry 6 in Table 1), whereas rotaxane **Mn3** exhibits about twice the catalytic activity when **tbpy** is bound (entry 8 in Table 1). This observation is explained by the fact that a pyridine axial ligand enhances the catalytic activity of Mn(III) porphyrins¹⁹, as also seen with **Mn2** ($\sim 2 \times$ higher, entry 7 in Table 1). With **Mn1**, when the axial ligand is present, catalytic conversion of the polymer substrate **bipy₂PD** takes place ‘inside’ the catalyst cavity, making the reaction sterically demanding; once the ligand is absent, reaction occurs more rapidly on the ‘outside’ of the catalyst. The polymer rotaxane **Mn3**, on the other hand, requires the **tbpy** ligand for activation and is considerably less active in the absence of **tbpy**. These results highlight that the **tbpy** ligand remains bound under the catalytic conditions used and that catalysis occurs inside the cavity of rotaxane **Mn3** (Fig. 1f).

Further support for the rotaxane mechanism was obtained from a study of the relative reactivities of the olefinic double bonds in the **bipy₂PD** substrate (also the thread in **Mn3**). In contrast to **PD**, this

polymer is a mixture of *cis*-1,4-butadiene, *trans*-1,4-butadiene and 1,2-butadiene units (Fig. 2) that differ in their reactivities towards oxidation by the Mn(III) porphyrin catalysts, with the *cis*-isomers generally being more reactive than the *trans*-isomers¹⁹. The ratios of the *trans*-epoxide/*cis*-epoxide products obtained with the different catalysts was measured by ¹H NMR after 30% total conversion of the polymeric substrate (Table 2). In the presence of **tbpy**, the highest *trans/cis* ratio was found for the rotaxane **Mn3** and the lowest for **Mn2** with **Mn1** displaying an intermediate ratio. All three catalysts showed a more similar product ratio in the absence of **tbpy** (*trans/cis* ratio = 1.0 ± 0.4). These results are again in line with the idea that in the case of **Mn1** and **Mn3**, catalysis occurs preferentially or completely within the cavity of the catalysts. To estimate the extent to which catalysis occurs inside and outside **Mn1**, we carried out control epoxidation reactions using a **Mn1** derivative, in which the cavity is completely blocked in a rotaxane architecture (**Mn4**, see Supplementary Information S-7). These experiments revealed that for the catalytic system **Mn1-PD**, 80% of the polymer conversion occurs inside **Mn1** and for the polymer system **Mn1-bipy₂PD** 45% inside. This latter value is in good agreement with the 55% conversion inside **Mn1**, obtained from the analysis of the product *cis/trans* ratios (see Supplementary Information S-7). This lower percentage conversion inside the cavity of **Mn1**, for the **Mn1-bipy₂PD** system is probably the result of the endgroups of the polymeric substrate binding and partially inhibiting catalysis inside the toroid.

It is unclear at this stage whether the catalytic process is sequentially processive or random, that is, whether the catalyst moves step by step along the chain or hops randomly. Calculated from the rate of the reaction and taking into account the length of the polymer, the catalyst would move around $1\text{--}700 \text{ pm s}^{-1}$ (the lower value if the reaction is sequential, the higher one if it is random) which is not very dissimilar to the speed of RNA polymerase ($3,000 \text{ pm s}^{-1}$)²⁶ but is low when compared to the speeds that can be calculated from data published on synthetic rotaxane shuttles (several $\mu\text{m s}^{-1}$)²⁷.

Experiments were also carried out with NaOCl as oxidant in a two-phase system of dichloromethane and water ($\text{CH}_2\text{Cl}_2/\text{H}_2\text{O}$). Using the conditions of entries 1–4 in Table 1, greatly enhanced activities were found for the cavity containing catalyst **Mn1** with turnover frequencies $>500 \text{ h}^{-1}$ for the conversion of **PD** to **PE**. This improved performance is thought to be a combination of the different nature of the oxidant used, and possibly the fact that the hydrophilic **PE** chain is pulled into the water phase as it forms during reaction, creating an additional driving force for the reaction (see Supplementary Fig. S2). Unfortunately, no decisive conclusions could be made about the mechanism in this two-phase system, because experiments with rotaxane **Mn3** were unsuccessful—the stoppers were cleaved off under the strongly basic reaction conditions (pH ~ 13)²⁸, destroying the rotaxane topology.

The results presented here show that catalytic modifications of a polymer can be realized using a catalyst that is topologically linked to its substrate. To make the system sequentially processive, one has to precisely balance the speed of the movement of the catalyst and the catalytic rate of the reaction, which in the present case is probably too slow when compared to the former process. Nevertheless, the rotaxanes **Mn1:PD** and **Mn3** are initial attempts to mimic the naturally occurring toroidal enzyme systems such as the various exo- and endonucleases. Processive rotaxane catalysis may open a new route, at least in principle, to carry out post-polymerization transformations on various functional polymers, either synthetic or biological. □

Methods

Catalytic activities were determined as follows: in a typical experiment the oxidant (1–2 equiv. PhIO) was added to a deuterated chloroform solution (1.0 ml) containing the porphyrin and polymer and when appropriate, the **tbpy** ($500 \times [\text{porphyrin}]$) ligand

Table 2 Chemo- and stereo-selectivities of reactions

Substrate	Catalyst	Polyepoxides formed		
		<i>cis</i>	<i>trans</i>	<i>trans/cis</i> ratio
PD*	Mn1	20%	80%	4.0
PD*	Mn2	78%	22%	0.3
bipy₂PD†	Mn1	37%	63%	1.7
bipy₂PD†	Mn2	59%	41%	0.7
Mn3‡	Mn3	20%	80%	4.0

All reactions were carried out at [C = C] of 6–8 mM in the presence of 500-fold excess of **tbpy**. See Table 1 for abbreviations and Methods for experimental details.

*This substrate (**PD**) contains $>98\%$ *cis*-butadiene bonds.

†These substrates (**bipy₂PD** and **Mn3**) contain 22% 1,4-*cis*-, 61% 1,4-*trans*- and 17% 1,2-butadiene double bonds (Fig. 2), giving a *trans/cis* ratio of 2.8 in the starting product.

under an argon atmosphere. For experiments with **Mn3**, a concentration of $[\text{Mn3}] = 0.08 \text{ mM}$ was used (on the basis of the relative molecular mass, M_w , of **Mn3** this gives $[\text{porphyrin}] = 0.16 \text{ mM}$ and $[\text{C} = \text{C}] = 6 \text{ mM}$) and when appropriate the ligand **tbpy** in a concentration of $[\text{tbpy}] = 81 \text{ mM}$. The resulting mixture was magnetically stirred at 1,100 r.p.m. Samples were taken which were filtered (to remove excess of iodosylbenzene) and the resulting filtrates were analysed by $^1\text{H-NMR}$. In the case of **Mn3** and **bipy2PD** only the protons of the *cis*- and *trans*-1,4-polybutadiene (5.8–5.2 p.p.m.) and *cis*- and *trans*-1,4-polyepoxide (3.0–2.6 p.p.m.) were used to calculate the conversion of the substrate. Turnover frequencies were calculated from data points up to ~30% conversion, since in a number of cases at higher conversions, deviations from first-order behaviour in substrate (alkene) concentration were observed. See Supplementary Information for further experimental details.

Received 19 February; accepted 17 July 2003; doi:10.1038/nature01925.

- Breyer, W. A. & Matthew, B. M. A structural basis for processivity. *Protein Sci.* **10**, 1699–1711 (2001).
- Kool, E. T., Morales, J. C. & Guckian, K. M. Mimicking the structure and function of DNA: Insights into DNA stability and replication. *Angew. Chem. Int. Edn Engl.* **39**, 990–1009 (2000).
- Benkovic, S. J., Valentine, A. M. & Salinas, F. Replisome-mediated DNA replication. *Annu. Rev. Biochem.* **70**, 181–208 (2001).
- Lin, S. M., Lloyd, R. S. & Roberts, R. J. *Nucleases*, 2nd edn (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1993).
- Kool, E. T. Recognition of DNA, RNA, and proteins by circular oligonucleotides. *Acc. Chem. Res.* **31**, 502–510 (1998).
- Trakselis, M. A., Alley, S. C., Abel-Santos, E. & Benkovic, S. J. Creating a dynamic picture of the sliding clamp during T4 DNA polymerase holoenzyme assembly by using fluorescence resonance energy transfer. *Proc. Natl Acad. Sci. USA* **98**, 8368–8375 (2001).
- Kovall, R. & Matthews, B. W. Toroidal structure of lambda-exonuclease. *Science* **277**, 1824–1827 (1997).
- Ashton, P. R. *et al.* A light-fueled “piston cylinder” molecular-level machine. *J. Am. Chem. Soc.* **120**, 11190–11191 (1998).
- Sauvage, J.-P. Transition metal-containing rotaxanes and catenanes in motion: towards molecular machines and motors. *Acc. Chem. Res.* **31**, 611–619 (1998).
- Davis, A. P. Synthetic molecular motors. *Nature* **401**, 120–121 (1999).
- Schalley, C. A., Beizai, K. & Vögtle, F. On the way to rotaxane-based molecular motors: studies in molecular mobility and topological chirality. *Acc. Chem. Res.* **34**, 465–476 (2001).
- Bissell, R. A., Córdova, E., Kaifer, A. E. & Stoddart, J. F. A chemically and electrochemically switchable molecular shuttle. *Nature* **369**, 133–137 (1994).
- Jiménez, M. C., Dietrich-Buchecker, C. & Sauvage, J.-P. Towards synthetic molecular muscles: Contraction and stretching of a linear rotaxane dimer. *Angew. Chem. Int. Edn Engl.* **39**, 3284–3287 (2000).
- Collier, C. P. *et al.* Electronically configurable molecular-based logic gates. *Science* **285**, 391–394 (1999).
- Harada, A., Li, J. & Kamachi, M. The molecular necklace: a rotaxane containing many threaded α -cyclodextrins. *Nature* **356**, 325–327 (1992).
- Shen, Y. X., Xie, D. & Gibson, H. W. Polyrotaxanes based on polyurethane backbones and crown ether cyclis. 1. Synthesis. *J. Am. Chem. Soc.* **116**, 537–548 (1994).
- Rowan, A. E., Aarts, P. P. M. & Koutstaal, K. W. M. Novel porphyrin-viologen rotaxanes. *Chem. Commun.*, 611–612 (1998).
- Elemans, J. A. A. W. *et al.* Porphyrin clips derived from diphenylglycoluril. Synthesis, conformational analysis, and binding properties. *J. Org. Chem.* **64**, 7009–7016 (1999).
- Meunier, B. Metalloporphyrins as versatile catalysts for oxidation reactions and oxidative DNA cleavage. *Chem. Rev.* **92**, 1411–1456 (1992).
- Elemans, J. A. A. W., Bijsterveld, E. J. A., Rowan, A. E. & Nolte, R. J. M. A host-guest epoxidation catalyst with enhanced activity and stability. *Chem. Commun.*, 2443–2444 (2000).
- Tornaritis, M. J. & Coutsolelos, A. G. Metalloporphyrins catalyse *cis*-polybutadiene to polyepoxide. *Polymer* **33**, 1771–1772 (1992).
- Sacco, H. C., Iamamoto, Y. & Lindsay Smith, J. R. Alkene epoxidation with iodosylbenzene catalysed by polyionic supports. *J. Chem. Soc. Perkin Trans. 2*, 181–190 (2001).
- Tsuda, Y., Takahashi, K., Yamaguchi, T., Matsui, S. & Komura, T. Catalytic epoxidation of cyclohexene by covalently linked manganese porphyrin-viologen complex. *J. Mol. Catal. A: Chem.* **130**, 285–295 (1998).
- Hollis, B. W. 25-Hydroxyvitamin D₃-1 α -hydroxylase in porcine hepatic tissue: Subcellular localization to both mitochondria and microsomes. *Proc. Natl Acad. Sci. USA* **87**, 6009–6013 (1990).
- Guengerich, F. P. Reactions and significance of cytochrome P-450 enzymes. *J. Biol. Chem.* **266**, 10019–10022 (1991).
- Perkins, T. T., Mitsis, P. G., Dalal, R. V. & Block, S. M. Watching enzymes move along DNA one at a time. *Biophys. J.* **80**, 1464 Part 2 (2001).
- Brouwer, A. M. *et al.* Photoinduction of fast, reversible translational motion in a hydrogen-bonded molecular shuttle. *Science* **291**, 2124–2128 (2001).
- Farrington, J. A., Ledwith, A. & Stam, M. F. Cation-radicals: Oxidation of methoxide ion with 1,1'-dimethyl-4,4'-bipyridylium dichloride (paraquat dichloride). *J. Chem. Soc. Chem. Commun.*, 259–260 (1969).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. Foekema and I. M. Dixon for preliminary studies and discussions. This research was supported by a NRSC Catalysis grant, a NWO Vidi grant (A.E.R.) and an EC Marie Curie fellowship (P.T.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.E.R. (rowan@sci.kun.nl).

Extreme deuterium enrichment in stratospheric hydrogen and the global atmospheric budget of H₂

Thom Rahn^{1*}, John M. Eiler¹, Kristie A. Boering^{3,4}, Paul O. Wennberg^{1,2}, Michael C. McCarthy³, Stanley Tyler⁵, Sue Schauffler⁶, Stephen Donnelly⁶ & Elliot Atlas⁶

¹Division of Geological and Planetary Sciences, and ²Division of Engineering and Applied Sciences, California Institute of Technology, Pasadena, California 91125, USA

³Department of Chemistry, and ⁴Department of Earth and Planetary Science, University of California at Berkeley, Berkeley, California 94720, USA

⁵Earth Systems Science Department, University of California, Irvine, Irvine, California 92697, USA

⁶National Center for Atmospheric Research, Boulder, Colorado 80307-3000, USA

* Present address: Los Alamos National Laboratory, EES-6, MS-D462, Los Alamos, New Mexico 87545, USA

Molecular hydrogen (H₂) is the second most abundant trace gas in the atmosphere after methane¹ (CH₄). In the troposphere, the D/H ratio of H₂ is enriched by 120‰ relative to the world's oceans^{2–4}. This cannot be explained by the sources of H₂ for which the D/H ratio has been measured to date (for example, fossil fuels and biomass burning)^{5,6}. But the isotopic composition of H₂ from its single largest source—the photochemical oxidation of methane—has yet to be determined. Here we show that the D/H ratio of stratospheric H₂ develops enrichments greater than 440‰, the most extreme D/H enrichment observed in a terrestrial material. We estimate the D/H ratio of H₂ produced from CH₄ in the stratosphere, where production is isolated from the influences of non-photochemical sources and sinks, showing that the chain of reactions producing H₂ from CH₄ concentrates D in the product H₂. This enrichment, which we estimate is similar on a global average in the troposphere, contributes substantially to the D/H ratio of tropospheric H₂.

H₂ has been proposed as the basis for fuel-cell technologies that are anticipated to expand substantially in coming decades, and, as with fossil fuels, there is potential for significant leakage from the requisite infrastructure^{7,8}. The consequences of a new source of H₂ to the atmosphere are not easily anticipated, and preparation for potential change must begin with a precise and accurate description of the global budget. Stable isotope measurements are often used to constrain the budgets of atmospheric trace gases, but have seen limited use for H₂ owing to analytical difficulties and lack of data on its isotopically distinct sources and sinks. Major sources having known isotopic compositions (fossil-fuel combustion, biomass burning, H₂-producing metabolisms) have $\delta\text{D}_{\text{H}_2}$ values ranging from –800‰ to –250‰ (refs 2–6), far lower than the average tropospheric value of about +120‰ (refs 2–4, 9). It has been argued that this difference could be caused by the slower rate of photochemical oxidation of HD compared to HH (if H₂ oxidation is the major tropospheric sink)¹⁰, or by enrichment of D in H₂ produced via oxidation of methane (CH₄) and/or non-methane hydrocarbons (if uptake by soils is the major tropospheric sink)⁵. Records of H₂ concentration¹¹ show greater seasonal amplitude and lower average concentration in the Northern Hemisphere, implying an atmospheric lifetime of ~2 yr and requiring uptake by soils to be the major sink¹. Thus, either photochemical production is indeed responsible for the observed D enrichment or our understanding of the H₂ budget is incomplete.

Oxidation of atmospheric CH₄ consists of a chain of reactions that include H₂ as one of the products (Fig. 1). Although roughly 90% of this photochemical CH₄ loss occurs in the troposphere, we

focus on a suite of stratospheric samples (see Methods) because the D content of stratospheric H_2 and CH_4 can be studied in relative isolation from non-photochemical sources and sinks of these gases. Mixing ratios of CH_4 (ref. 12) and H_2 are plotted in Fig. 2 as a function of nitrous oxide (N_2O) mixing ratios¹³, a proxy for stratospheric mean age¹⁴. Methane decreases with increasing age from 1,750 to 700 p.p.b. over ~ 6 yr in the stratosphere (reflecting its photochemical destruction), whereas H_2 remains at tropospheric levels of ~ 525 p.p.b. This relatively constant level of H_2 has been observed previously¹⁵, and indicates the approximate balance of loss due to photochemical oxidation by OH, Cl and $\text{O}(^1\text{D})$ and production by CH_4 oxidation¹⁵. Figure 3 shows $\delta\text{D}_{\text{H}_2}$ as a function of CH_4 mixing ratio. The maximum $\delta\text{D}_{\text{H}_2}$ value measured ($+440\text{‰}$) is hundreds of per mil enriched in D compared to any other terrestrial material; the only other natural materials found on Earth having comparable δD values are carbonaceous chondrites¹⁶ and the SNC meteorites¹⁷.

The observed increase in $\delta\text{D}_{\text{H}_2}$ with stratospheric mean age reflects the effects of two processes: (1) preferential removal of HH relative to HD from the H_2 pool during H_2 oxidation, and (2) preferential addition of HD relative to HH to the H_2 pool during the chain of reactions initiated by methane oxidation. In order to assess the magnitudes of these different effects, we examine the reactions responsible for both H_2 loss and production. These reactions are itemized in Table 1; R_i denotes a reaction involving D-free isotopologues (for example, HH or CH_4), and R_i' a reaction involving

D-bearing isotopologues (for example, HD or CH_3D). The rate coefficients are denoted by k_i and k_i' , respectively, and their ratios are the fractionation factors, $\alpha_i = k_i'/k_i$. As each of the reactions R1 to R3' and R4 to R6' contribute to the oxidation of hydrogen or methane, respectively, we also define the oxidant-weighted fractionation factors $\alpha_{\text{H}_2+\text{ox}}$ and $\alpha_{\text{CH}_4+\text{ox}}$ as:

$$\alpha_{\text{H}_2+\text{ox}} = (k_1'/k_1)f_{\text{OH}} + (k_2'/k_2)f_{\text{O}(^1\text{D})} + (k_3'/k_3)f_{\text{Cl}} \quad (1)$$

where f is the fraction of H_2 (or CH_4) loss due to each oxidant; a similar expression for $\alpha_{\text{CH}_4+\text{ox}}$ involves substituting reactions R4–R6' for R1–R3'. In addition to the fractionation due to different reaction rates (for example, k_1'/k_1), branching in R4', R5' and R6' is an additional source of isotopic fractionation (Fig. 1). For instance, abstraction of H (or D) from CH_3D in R4' by OH can yield the products $\text{CH}_2\text{D} + \text{H}_2\text{O}$ (or $\text{CH}_3 + \text{HDO}$). Fractionation also occurs when the methyl radical (CH_3 or CH_2D) product of reactions R4–R6' is oxidized (Fig. 1), ultimately (and quantitatively) yielding formaldehyde (HCHO or CHDO). Formaldehyde then reacts with OH, $\text{O}(^1\text{D})$ or Cl (R7) or is photolysed (R8a, b) within several hours. The proportion of HCHO that is photolysed and the quantum yield for the two channels (R8a and R8b) both depend on pressure, temperature, and wavelength of light¹⁸, and thus vary with altitude and latitude. We define the product of these subsequent fractionations with the initial $\alpha_{\text{CH}_4+\text{ox}}$ as the total fractionation, $\bar{\alpha}_{\text{CH}_4\rightarrow\text{H}_2} = 2(\gamma'/\gamma)\alpha_{\text{CH}_4+\text{ox}}$ where γ' and γ are the fractional yield of HD from CH_3D and HH from CH_4 , respectively (see Methods).

In order to obtain an estimate of $\bar{\alpha}_{\text{CH}_4\rightarrow\text{H}_2}$, we use a simple box model (see Methods) with the following assumptions. First, because the photochemical lifetime of CH_4 in the stratosphere decreases from almost 100 yr at the tropopause to less than one year in the upper stratosphere¹⁹, a characteristic altitude and latitude of 30 km in the tropics, where CH_4 loss rates are among the largest, is chosen. Second, the fractionation during H_2 loss is approximated by combining estimates of number densities for OH, Cl and $\text{O}(^1\text{D})$ (ref. 20) and rate coefficients for R1–R3' (refs 21–23) at 230 K, yielding a value for $\alpha_{\text{H}_2+\text{ox}}$ of 0.750. Third, we note that γ , the yield of H_2 for each CH_4 lost, depends on altitude and latitude and varies from 0.65 near the tropopause to 0.3 in the upper stratosphere (M.C.M., P. Connell and K.A.B., manuscript in preparation). At a characteristic altitude of 30 km, γ is ~ 0.5 . From these parameters we calculate a best-fit value of $\bar{\alpha}_{\text{CH}_4\rightarrow\text{H}_2} = 1.33$ from equation (4).

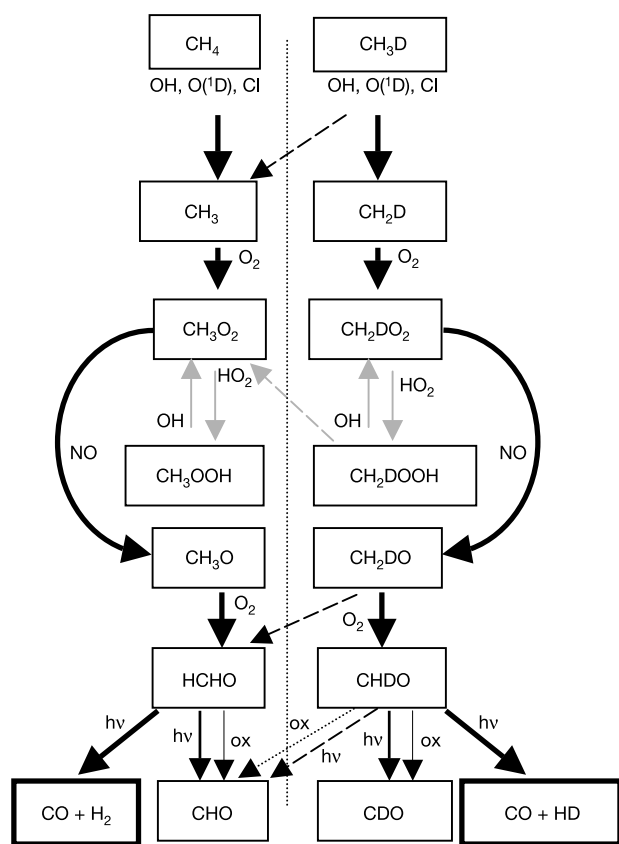


Figure 1 Simplified schematic diagram of the reaction pathway from methane to molecular hydrogen for both normal CH_4 and isotopically substituted CH_3D . Reactions with multiple pathways have relative magnitudes represented by arrow thickness. Steps represented by grey are examples of additional reactions that are important in the troposphere under certain conditions, but are less significant in the stratosphere. Branching or abstraction reactions that remove D during the oxidation sequence do so irreversibly with respect to formation of HD, and are represented by the arrows crossing the centre line.

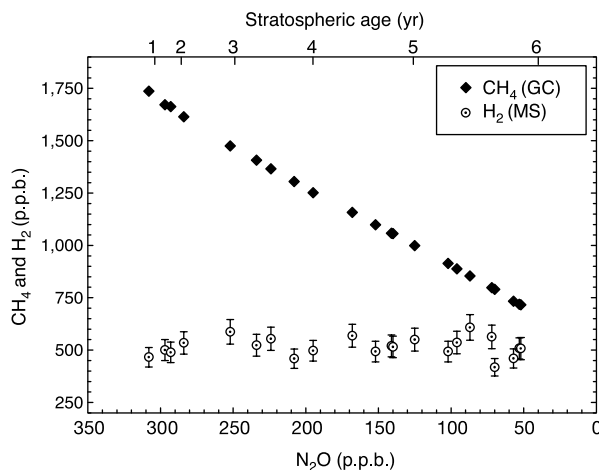


Figure 2 Methane and hydrogen mixing ratios in the stratosphere relative to nitrous oxide and mean age. Mean age is calculated from N_2O as a second-order polynomial fitted to the N_2O –age relationship from ref. 14 and originally derived from CO_2 observations. The error bars of the H_2 determined from the mass spectrometric results represent our estimated 2σ error.

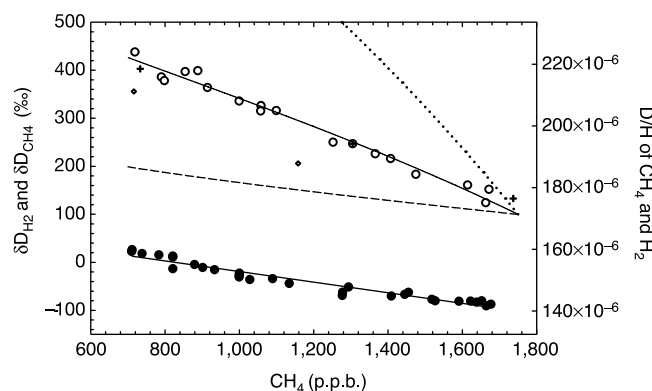


Figure 3 Plot of δD_{H_2} (open symbols) and δD_{CH_4} (filled circles) in the stratosphere against methane mixing ratios. (D content is also expressed as D/H ratio on the right ordinate.) Crosses and diamonds represent samples for which both the original canisters and subsequently filled glass flasks were analysed for δD_{H_2} , with diamonds representing those samples with suspected in-growth of H_2 as discussed in Supplementary Information. Solid line with H_2 data is modelled δD_{H_2} with $\bar{\alpha}_{CH_4 \rightarrow H_2} = 1.33$. Dotted line, expected trend if all D in CH_4 were conserved ($\bar{\alpha}_{CH_4 \rightarrow H_2} = 2.00$); dashed line, expected trend if D were conserved on a strictly statistical basis (that is, half of the D conserved, $\bar{\alpha}_{CH_4 \rightarrow H_2} = 1.00$).

This value shows that the fraction of $(D/H)_{H_2}$ produced for a given initial $(D/H)_{CH_4}$ is significantly greater than that of the original methane; for example, CH_4 entering the stratosphere with δD_{CH_4} of -90‰ would yield product H_2 with $\delta D_{H_2} = +213\text{‰}$.

Although a two- or three-dimensional model will be required to investigate the range of δD_{H_2} produced from CH_4 oxidation and its global average, the sensitivity of the box model to natural ranges for γ and α_{H_2+ox} provides conservative constraints on the value of $\bar{\alpha}_{CH_4 \rightarrow H_2}$ derived here. The integrated value of γ relevant to the stratosphere should be less than the maximum value of 0.65 at the tropopause and considerably larger than the minimum of 0.3 at the stratopause, so we vary γ as 0.5 ± 0.1 . Variability in α_{H_2+ox} due to changes in oxidant concentrations throughout the stratosphere is bounded by the maximum and minimum fractionations of the individual reactions^{21–23} (Table 1, R1–R3'). As it is highly unlikely that either of these limits is reached anywhere in the stratosphere, we vary α_{H_2+ox} as 0.75 ± 0.15 . These variations in γ and α_{H_2+ox} yield

a conservative range in $\bar{\alpha}_{CH_4 \rightarrow H_2}$ of $1.33^{+0.29}_{-0.25}$. Thus, $\bar{\alpha}_{CH_4 \rightarrow H_2}$ is expected to be greater than 1.0 everywhere in the stratosphere, showing that δD_{H_2} produced from methane oxidation is always enriched relative to the initial CH_4 . As a test for understanding stratospheric D enrichment, we also note that setting α_{H_2+ox} to unity (no fractionation due to the sink reactions with OH and Cl) in the model decreases the observed enrichment by 63%. The remaining enrichment results from photochemical production of HD, $\sim 25\%$ due to $\bar{\alpha}_{CH_4 \rightarrow H_2}$ and the remainder due to the progressive D enrichment of the source CH_4 .

Several arguments suggest that the value of $\bar{\alpha}_{CH_4 \rightarrow H_2}$ we estimate from stratospheric observations is also a reasonable approximation for the troposphere. A key variable controlling the isotopic composition of H_2 produced from CH_4 is α_{CH_4+ox} (larger values for α_{CH_4+ox} result in greater D enrichment of formaldehyde, which in turn produces greater D enrichment in H_2). We estimate, based on the negligible role of reaction with $O(^1D)$ and the temperature dependence of R4 and R4', that $\alpha_{CH_4+ox(trop)}$ in the troposphere is typically about 0.76, similar to $\alpha_{CH_4+ox(strat)} = 0.78$ calculated for the stratosphere (Table 1). Although additional reactions in the troposphere add complexity to the pathway of H_2 production (Fig. 1), we suggest that, because the fractionation during the initial oxidation of CH_4 is rate limiting, the above analysis will hold true.

Combining published estimates of the different tropospheric source terms¹ and the D content of those sources that are known^{5,6}, and making reasonable assumptions for those that have not been measured (for example, oceanic and nitrogen fixation sources), we use our value for $\bar{\alpha}_{CH_4 \rightarrow H_2}$ to estimate the relative strengths of the two most important H_2 sinks, reaction with OH and uptake in soils. From our isotopic analysis, we estimate that $\sim 77\%$ of tropospheric H_2 loss is due to soil uptake, in good agreement with the 75% estimated by Novelli *et al.*¹ and the 80% estimated by Hauglustaine and Ehrl²⁴ from concentration constraints alone. If we include stratosphere/troposphere exchange with an average δD_{H_2} of 200‰ for H_2 returning from the stratosphere (assuming a stratospheric turnover time of 2.5 yr; ref. 25), the soil sink proportion increases to $\sim 79\%$. We note that although the uncertainty in our budget estimate is large (about $\pm 20\%$) given the conservative box-model estimates for $\bar{\alpha}_{CH_4 \rightarrow H_2}$, our observations show that the enigma of heavy tropospheric H_2 can indeed be understood to result from HD production via CH_4 oxidation, as proposed by Gerst and Quay⁵. Although our estimate of $\bar{\alpha}_{CH_4 \rightarrow H_2}$ yields δD_{H_2} greater than the $130 \pm 70\text{‰}$ required for the photo-

Table 1 Reactions causing H_2 loss or production

R/	Reaction	f^*	α at 230 K†	α at 280 K†
R1	$H_2 + OH \rightarrow H_2O + H$	0.50		
R1'	$HD + OH \rightarrow HDO + H$ or $H_2O + D$		0.51 ²¹	0.56 ²¹
R2	$H_2 + O(^1D) \rightarrow OH + H$	0.49		
R2'	$HD + O(^1D) \rightarrow OD + H$ or $OH + D$		1.0 ²²	1.0 ²²
R3	$H_2 + Cl \rightarrow HCl + H$	0.01		
R3'	$HD + Cl \rightarrow HCl + D$ or $DCl + H$		0.52 ²³	0.44 ²³
R4	$CH_4 + OH \rightarrow H_2O + CH_3$	0.42		
R4'	$CH_3D + OH \rightarrow HDO + CH_3$ or $H_2O + CH_2D$		0.67 ²⁷	0.76 ²⁷
R5	$CH_4 + O(^1D) \rightarrow OH + CH_3$	0.43		
R5'	$CH_3D + O(^1D) \rightarrow OD + CH_3$ or $OH + CH_2D$		0.94 ²⁷	0.94 ²⁷
R6	$CH_4 + Cl \rightarrow HCl + CH_3$	0.15		
R6'	$CH_3D + Cl \rightarrow DCl + CH_3$ or $HCl + CH_2D$		0.63 ²⁸	0.65 ²⁸
R7	$HCHO + OH$ (or $O(^1D)$ or Cl) $\rightarrow$ products			
R7'	$HCDO + OH$ (or $O(^1D)$ or Cl) $\rightarrow$ products			
R8a	$HCHO + h\nu \rightarrow H_2 + CO$		Unknown	Unknown
R8a'	$HCDO + h\nu \rightarrow HD + CO$		Unknown	Unknown
R8b	$HCHO + h\nu \rightarrow H + HCO$			
R8b'	$HCDO + h\nu \rightarrow D + HCO$ or $H + DCO$		Unknown	Unknown

*Fractional loss of H_2 or CH_4 attributed to the respective oxidants at 30 km.

†Fractionations for the individual reactions at a characteristic stratospheric (230 K) and tropospheric (280 K) temperature.

‡Oxidant-weighted fractionation factor for H_2 used in the box model.

§Oxidant-weighted fractionation factor for CH_4 in the stratosphere for comparison with R4 and R4' at tropospheric temperature. This value shows more fractionation than the observed stratospheric average value of 0.87, which is attenuated by transport and mixing²⁶.

chemical source in the proposed Gerst and Quay isotopic budget, we note that our budget assumes thermodynamic equilibrium between H_2 and H_2O for the highly depleted oceanic and N_2 fixation sources ($\delta\text{D}_{\text{H}_2} \approx -700\text{‰}$), whereas Gerst and Quay included these terms with photochemical production. If these minor sources are indeed extremely depleted, as observed for other biogenic sources⁶, they will have considerable effect on the isotopic budget. Future research will need to address details such as these to improve upon the isotopic budget estimates. □

Methods

Isotopic notation

$\delta\text{D} = (R_{\text{sample}}/R_{\text{reference}} - 1) \times 1,000$, where R_i is the D/H ratio and the reference is Vienna Standard Mean Ocean Water.

Sampling

Samples were collected from the NASA ER-2 aircraft¹³ in January–March 2000 between 65–80°N, 13–63°E and 11–21 km altitude during the Sage III Ozone Loss and Validation Experiment²⁶. Duplicate analyses of selected samples that were transferred to glass flasks soon after collection showed, with only two exceptions, that H_2 and (D/H) H_2 were not affected by long-term storage in the electropolished canisters (Supplementary Information).

Box model derivation

Because $[\text{H}_2]$ is essentially constant (Fig. 2) in the stratosphere, the production rate of H_2 (P_{H_2}) is approximately equal to its loss rate¹⁵, L_{H_2} , which can be expressed as:

$$\frac{d[\text{H}_2]}{dt} = P_{\text{H}_2} - L_{\text{H}_2} \approx 0 \quad (2)$$

Note that P_{H_2} depends on the loss rate of CH_4 (L_{CH_4}) because H_2 is a by-product of the photooxidation of methane (R4–R7b; Fig. 3). As only a fraction of methane loss ultimately yields H_2 , we define this dependence as $P_{\text{H}_2} = \gamma L_{\text{CH}_4}$ where γ is the fractional yield of H_2 from CH_4 . The production of HD is similarly defined as $P_{\text{HD}} = \gamma' L_{\text{CH}_3\text{D}}$. The fraction of CH_3D that produces HD, γ' , is not equal to γ owing to the branching and subsequent fractionations that CH_3D and its products undergo during production of HD (see Fig. 1). These terms are unknown, and defined here as the composite term α_β . We define γ' as the product of γ with this composite fractionation so that it can be rewritten as $\alpha_\beta = \gamma'/\gamma$. Then, given that $[\text{H}_2] \approx [\text{HH}]$ and that the loss rates of HD and CH_3D can be expressed as $L_{\text{HD}} = L_{\text{HH}}\alpha_{\text{H}_2+\text{ox}}[\text{HD}]/[\text{HH}]$ and $L_{\text{CH}_3\text{D}} = L_{\text{CH}_4}\alpha_{\text{CH}_4+\text{ox}}[\text{CH}_3\text{D}]/[\text{CH}_4]$, respectively, and recalling that $P_{\text{HH}} = \gamma L_{\text{CH}_4}$, we derive an expression for the rate of change of HD:

$$\frac{d[\text{HD}]}{dt} = \gamma L_{\text{CH}_4} \frac{[\text{CH}_3\text{D}]}{[\text{CH}_4]} \alpha_\beta \alpha_{\text{CH}_4+\text{ox}} - L_{\text{HH}} \frac{[\text{HD}]}{[\text{HH}]} \alpha_{\text{H}_2+\text{ox}} \quad (3)$$

Because we express our data in terms of the (D/H) ratio rather than abundances of isotopologues, we note that $(\text{D}/\text{H})_{\text{H}_2} \approx 0.5[\text{HD}]/[\text{HH}]$ and $(\text{D}/\text{H})_{\text{CH}_4} \approx 0.25[\text{CH}_3\text{D}]/[\text{CH}_4]$ and define the total fractionation in production of H_2 from CH_4 as $\bar{\alpha}_{\text{CH}_4 \rightarrow \text{H}_2} = 2\alpha_\beta \alpha_{\text{CH}_4+\text{ox}} = 2(\gamma'/\gamma) \alpha_{\text{CH}_4+\text{ox}}$. It is this fractionation that determines the D content of H_2 produced from methane oxidation, and can be estimated from our stratospheric measurements of δD of H_2 and CH_4 . Included in $\bar{\alpha}_{\text{CH}_4 \rightarrow \text{H}_2}$ is the initial oxidation of CH_4 , $\alpha_{\text{CH}_4+\text{ox}}$, as well as all subsequent branching and fractionation, α_β , in the formation of H_2 . Using the relationships above, and recalling that we assume $[\text{HH}]$ to be constant, equation (3) can be rewritten as:

$$\frac{d(\text{D}/\text{H})_{\text{H}_2}}{dt} = [\text{HH}]^{-1} [\gamma L_{\text{CH}_4} (\text{D}/\text{H})_{\text{CH}_4} \bar{\alpha}_{\text{CH}_4 \rightarrow \text{H}_2} - L_{\text{HH}} (\text{D}/\text{H})_{\text{H}_2} \alpha_{\text{H}_2+\text{ox}}] \quad (4)$$

We then combine equation (4) with the measurements of $(\text{D}/\text{H})_{\text{H}_2}$ reported here and of $(\text{D}/\text{H})_{\text{CH}_4}$ for the same suite of samples¹² along with the assumptions described in the text in order to solve for $\bar{\alpha}_{\text{CH}_4 \rightarrow \text{H}_2}$.

Received 24 January; accepted 15 July 2003; doi:10.1038/nature01917.

- Novelli, P. C. *et al.* Molecular hydrogen in the troposphere: Global distribution and budget. *J. Geophys. Res.* **104**, 30427–30444 (1999).
- Ehhalt, D., Israel, G., Roether, W. & Stich, W. Tritium and deuterium content of atmospheric hydrogen. *J. Geophys. Res.* **68**, 3747–3751 (1963).
- Gonsior, B., Friedman, I. & Lindenmayr, G. New tritium and deuterium measurements in atmospheric hydrogen. *Tellus XVIII*, 256–261 (1966).
- Friedman, I. & Scholz, T. G. Isotopic composition of atmospheric hydrogen, 1967–1969. *J. Geophys. Res.* **79**, 785–788 (1974).
- Gerst, S. & Quay, P. Deuterium component of the global molecular hydrogen cycle. *J. Geophys. Res.* **106**, 5021–5031 (2001).
- Rahn, T., Kitchen, N. & Eiler, J. M. D/H ratios of atmospheric H_2 in urban air: Results using new methods for analysis of nano-molar H_2 samples. *Geochim. Cosmochim. Acta* **66**, 2475–2481 (2002).
- Blake, D. R. & Rowland, F. S. Urban leakage of liquefied petroleum gas and its impact on Mexico City air quality. *Science* **269**, 953–956 (1995).
- Chen, T. Y., Simpson, I. J., Blake, D. R. & Rowland, F. S. Impact of the leakage of liquefied petroleum gas (LPG) on Santiago air quality. *Geophys. Res. Lett.* **28**, 2193–2196 (2001).
- Gerst, S. & Quay, P. The deuterium content of atmospheric molecular hydrogen: Method and initial measurements. *J. Geophys. Res.* **105**, 26433–26445 (2000).
- Ehhalt, D. H., Davidson, J. A., Cantrell, C. A., Friedman, I. & Tyler, S. The kinetic isotope effect in the reaction of H_2 with OH. *J. Geophys. Res.* **94**, 9831–9836 (1989).

- Khalil, M. A. K. & Rasmussen, R. A. Global increase of atmospheric molecular hydrogen. *Nature* **347**, 743–745 (1990).
- Rice, A. L., Tyler, S., McCarthy, M. C., Boering, K. A. & Atlas, E. The carbon and hydrogen isotopic composition of stratospheric methane: Part 1. High-precision observations from the NASA ER-2 aircraft. *J. Geophys. Res.* doi:10.1029/2002JD003042 (in the press).
- Schaffler, S. M. *et al.* Chlorine budget and partitioning during the Stratospheric Aerosol and Gas Experiment (SAGE) III Ozone Loss and Validation Experiment (SOLVE). *J. Geophys. Res.* **108**, doi:10.1029/2001JD002040 (2003).
- Boering, K. A. *et al.* Stratospheric mean ages and transport rates from observations of carbon dioxide and nitrous oxide. *Science* **274**, 1340–1343 (1996).
- Hurst, D. F. *et al.* Closure of the total hydrogen budget of the northern extratropical lower stratosphere. *J. Geophys. Res.* **104**, 8191–8200 (1999).
- Robert, F. & Epstein, S. The concentration and isotopic composition of hydrogen, carbon, and nitrogen in carbonaceous meteorites. *Geochim. Cosmochim. Acta* **46**, 81–95 (1982).
- Leshin, L. A., Epstein, S. & Stolper, E. M. Hydrogen isotope geochemistry of SNC meteorites. *Geochim. Cosmochim. Acta* **60**, 2635–2650 (1996).
- Moortgat, G., Seiler, W. & Warneck, P. Photodissociation of HCHO in air—CO and H_2 quantum yields at 220 and 300 K. *J. Chem. Phys.* **78**, 1185–1190 (1983).
- Brasseur, G. & Solomon, S. *Aeronomy of the Middle Atmosphere* (Reidel, Dordrecht, 1984).
- DeMore, W. B. *et al.* *Chemical Kinetics and Photochemical Data for Use in Stratospheric Modeling* 278 (NASA-JPL, Pasadena, 1997).
- Talukdar, R. K. *et al.* Kinetics of hydroxyl radical reactions with isotopically labeled hydrogen. *J. Phys. Chem.* **100**, 3037–3043 (1996).
- Talukdar, R. K. & Ravishankara, A. R. Rate coefficients for $\text{O}(^1\text{D}) + \text{H}_2$, D_2 , HD reactions and H atom yield in $\text{O}(^1\text{D}) + \text{HD}$ reaction. *Chem. Phys. Lett.* **253**, 177–183 (1996).
- Taatjes, C. A. Infrared frequency-modulation measurements of absolute rate coefficients for $\text{Cl} + \text{HD} \rightarrow \text{HCl}(\text{DCl}) + \text{D}(\text{H})$ between 295 and 700 K. *Chem. Phys. Lett.* **306**, 33–40 (1999).
- Hauglustaine, D. A. & Ehhalt, D. H. A three-dimensional model of molecular hydrogen in the troposphere. *J. Geophys. Res.* **107**, doi:10.1029/2001JD001156 (2002).
- Holton, J. R. On the global exchange of mass between the stratosphere and troposphere. *J. Atmos. Sci.* **47**, 392–395 (1990).
- Newman, P. A. *et al.* An overview of the SOLVE/THESEO 2000 campaign. *J. Geophys. Res.* **107**, doi:10.1029/2001JD001303 (2002).
- Saueressig, G. *et al.* Carbon 13 and D kinetic isotope effects in the reactions of CH_4 with $\text{O}(^1\text{D})$ and OH: New laboratory measurements and their implications for the isotopic composition of stratospheric methane. *J. Geophys. Res.* **106**, 23127–23138 (2001).
- Saueressig, G., Bergamaschi, P., Crowley, J. N., Fischer, H. & Harris, G. W. D/H kinetic isotope effect in the reaction $\text{CH}_4 + \text{Cl}$. *Geophys. Res. Lett.* **23**, 3619–3622 (1996).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank N. Kitchen for assistance in the laboratory. This study was supported in part by the NSF (T.A.R., K.A.B. and S.T.), the Davidow Fund and General Motors Corp. (J.M.E.), the David and Lucile Packard Foundation (K.A.B.) and the NASA Upper Atmosphere Research Program (P.O.W., K.A.B., E.A., S.S. and S.D.). The National Centre for Atmospheric Research is operated by the University Corporation for Atmospheric Research under the sponsorship of the NSF. T.R. is the recipient of a Frederick Reines postdoctoral fellowship in experimental sciences awarded by Los Alamos National Laboratory.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.R. (trahn@lanl.gov).

Remote triggering of deep earthquakes in the 2002 Tonga sequences

Rigobert Tibi¹, Douglas A. Wiens¹ & Hiroshi Inoue²

¹Department of Earth and Planetary Sciences, Washington University, St. Louis, Missouri 63130, USA

²National Research Institute for Earth Science and Disaster Prevention, Tsukuba, Ibaraki 305-0006, Japan

It is well established that an earthquake in the Earth's crust can trigger subsequent earthquakes, but such triggering has not been documented for deeper earthquakes. Models for shallow fault interactions suggest that static (permanent) stress changes can trigger nearby earthquakes, within a few fault lengths from the causative earthquake^{1–3}, whereas dynamic (transient) stresses carried by seismic waves may trigger earthquakes both nearby

and at remote distances^{4–8}. Here we present a detailed analysis of the 19 August 2002 Tonga deep earthquake sequences and show evidence for both static and dynamic triggering. Seven minutes after a magnitude 7.6 earthquake occurred at a depth of 598 km, a magnitude 7.7 earthquake (664 km depth) occurred 300 km away, in a previously aseismic region. We found that nearby aftershocks of the first mainshock are preferentially located in regions where static stresses are predicted to have been enhanced by the mainshock. But the second mainshock and other triggered events are located at larger distances where static stress increases should be negligible, thus suggesting dynamic triggering. The origin times of the triggered events do not correspond to arrival times of the main seismic waves from the mainshocks and the dynamically triggered earthquakes frequently occur in aseismic regions below or adjacent to the seismic zone. We propose that these events are triggered by transient effects in regions near criticality, but where earthquakes have difficulty nucleating without external influences.

Although earthquake triggering is a well-studied phenomenon for shallow events, there has been little work on triggered deep earthquakes. Understanding the conditions under which deep earthquakes may be triggered may provide important clues about the mechanism of deep earthquakes, which is still uncertain^{9–13}.

On 19 August 2002, a deep earthquake of moment magnitude M_w 7.6 in the Tonga subduction zone was followed 130 seconds later by an event of body-wave magnitude M_b 5.9 located 290 km away, and an additional 311 seconds later by an M_w 7.7 earthquake near that location (Fig. 1, Table 1). For the region within about 100 km from the later earthquakes, no well-located events are found in systematic relocations (Fig. 2), suggesting that these earthquakes occurred in a previously aseismic region. Therefore, they do not represent a fortuitous occurrence of background seismicity, and their proximity in time clearly indicates a causal link between the initial event and the later earthquakes. Thus the second main event and its foreshock must have been triggered by processes following the first earthquake. The second main event, located at a depth of 664 ± 4 km is also the deepest-known event with $M_w > 7$.

Rupture details for the two main events, and aftershock locations need to be clarified to understand the triggering scenario. We combined teleseismic arrival times with regional data to constrain the location of both sequences (see Methods). One foreshock and 21 aftershocks are reported for the first sequence. The second sequence consists of the foreshock mentioned above and three aftershocks (see Supplementary Table 2). Results of body waveform inversion (see Methods) suggest that the first main event consists of three episodes of moment release (Supplementary Table 3 and Supplementary Fig. 5). Rupture propagated mainly northward, consistent with the location of most aftershocks (Fig. 1a). However, some events are also located south of the main-event initiation

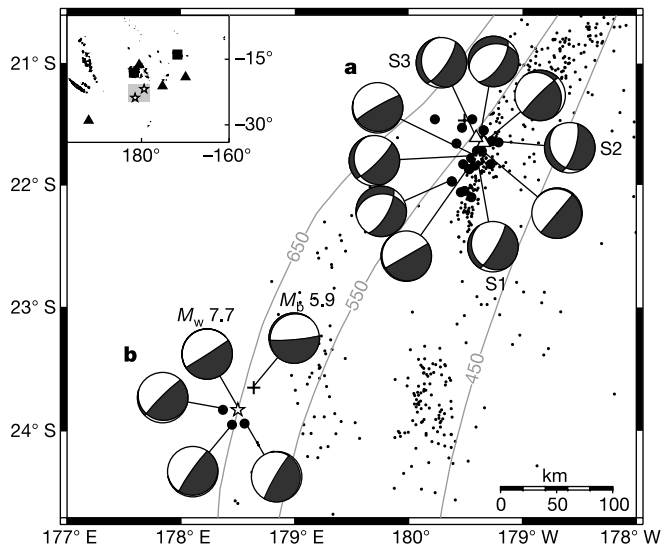


Figure 1 Epicentral locations of the two 2002 Tonga deep earthquake sequences. Dots, the locations of previous earthquakes³³. Grey lines, contours of deep seismicity³⁴ with the numbers indicating the depth in km to the seismogenic zone. Fault plane solutions are lower-hemisphere projections of the focal spheres. Black and white quadrants indicate compressional and tensional areas, respectively. **a** Star, the epicentre (S1) of the initial earthquake (M_w 7.6, depth 598 km), triangles, its second (S2) and third (S3) episodes of rupture, as inferred from the body-wave inversion (see Methods and Supplementary Table 3). Cross, a foreshock that occurred 8 days before the main event. Circles, aftershocks (Supplementary Table 2). The distribution of aftershocks excluding events located south of the mainshock epicentre (discussed in text) indicates an east–west and vertical rupture extent of about 30 km and 40 km, respectively. **b** Locations of the triggered earthquake (star) (M_w 7.7, depth 664 km), and its aftershocks (circles). Cross, a foreshock that occurred about five minutes before the main event. Inset, the locations of the SPANET (triangles) and the regional stations AFI and MSVF of the Global Seismic Network (squares) used in combination with teleseismic stations to locate both earthquake sequences. The SPANET network consists of broadband seismic stations operated in several islands in the Southwest Pacific. Stars in the inset, the epicentres of the 2002 mainshocks.

point. Results of stress-transfer calculations (discussed later) suggest that events located south from the mainshock initiation point are related to a southward expansion of the aftershock activity, resulting from Coulomb stress increase in that area (Fig. 3). The overall mechanism for the mainshock consists of a north–south striking and steeply dipping nodal plane, and a northwest–southeast trending subhorizontal nodal plane. Given the east–west rupture extent of only ~ 30 km (Fig. 1), the vertical component of rupture of about 40 km cannot be accommodated along the subhorizontal

Table 1 Triggered deep earthquakes

Initial event		Triggered Event		Mag.	Depth (km)	ΔT (min:s)	Dist. (km)	v_{st} (km s ⁻¹)	Notes
Date	OT (h:min:s)	Date	OT (h:min:s)						
19 Aug 2002	11:01:04			7.6	598				Tonga region
		19 Aug 2002	11:03:14	5.9	647	2:10	290	2.23	Nearly aseismic area
		19 Aug 2002	11:08:25	7.7	664	7:21	313	0.71	
09 Mar 1994	23:28:08			7.6	563				Tonga region
		10 Mar 1994	00:43:08	5.2	609	75:00	105	0.02	In slab
		10 Mar 1994	01:51:41	5.0	637	143:33	72	0.01	Nearly aseismic area
09 Jun 1994	00:33:17			8.3	640				Bolivia region
		09 Jun 1994	01:15:18	5.9	650	42:01	84	0.03	Aseismic region
		26 May 1986	19:06:16	7.1	543	25:31	257	0.17	Tonga region
26 May 1986	18:40:45			5.4	633	67:52	317	0.08	Nearly aseismic area
		26 May 1986	19:48:37	7.3	473				
		21 Jul 1994	18:55:59	5.9	514	19:26	127	0.11	Japan Sea
21 Jul 1994	18:36:33								In slab

OT, event origin time. ΔT , the separation in origin time. Dist., the epicentral distance between the causative earthquake and the triggered event. v_{st} , the estimated apparent stress pulse velocity.

nodal plane. An inversion for the plane that best fits the aftershock locations, excluding the triggered events (discussed later), yields values of 14° for the strike and 73° for the dip, consistent with the orientation of the near-vertical nodal plane. These represent strong evidence that rupture during the first 2002 main event occurred along the near vertical plane.

For the second mainshock, both visual inspection and inversion of the waveforms show no evidence of rupture directivity. The spatial distribution of the foreshock and aftershocks do not define a consistent direction relative to the mainshock. Thus the rupture may have spread out radially, starting from the initiation point (Fig. 1b), which would explain the absence of a directivity effect. The mainshock focal mechanism shows a northeastward trending vertical nodal plane and a southwestward striking horizontal nodal plane. The three aftershocks and the main event are located approximately at the same depth of ~ 665 km (Fig. 2). The lateral extent, as defined by the foreshock and aftershocks, is inconsistent with the orientation of the vertical nodal plane, suggesting that the horizontal plane of the focal mechanism is probably the one that ruptured.

The Coulomb failure criterion is broadly used to describe the conditions under which shear failure occurs in rocks¹⁴. For shallow events, Coulomb stress increases associated with prior earthquakes have successfully explained subsequent rupture (that is triggering) on nearby and remote faults, and increases in seismicity rates^{1–7}. We investigated fault interactions in the 2002 Tonga deep earthquakes (see Methods section). First, we examined the relationship between the first main event and its aftershocks. The results show that aftershocks south of the mainshock initiation point are located within and around an area where the Coulomb stress increased by about 0.2–3 MPa during the main rupture (Fig. 3). This suggests that the southern aftershocks were promoted by stress transfer from the mainshock, which may itself have been induced by the foreshock

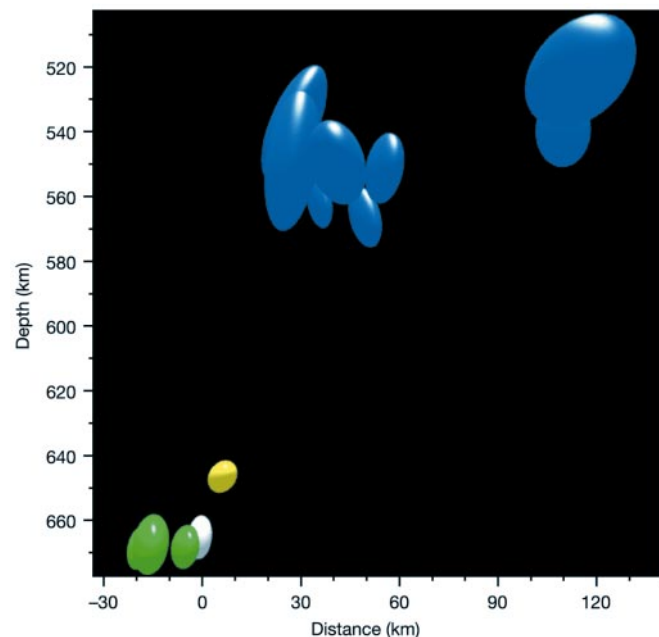


Figure 2 Vertical cross-section perpendicular to the deep Tonga slab showing the 95% confidence ellipsoid for the location of the triggered 2002 Tonga sequence. White, the main event; yellow, its foreshock; green, the aftershocks; blue, the relocated background seismicity from the latitude range between -23.25° and -24.25° , as reported in the International Seismological Centre (ISC) (1964–1989) and PDE (post-1989) catalogues. Note the lack of seismicity below about 570 km, indicating that the 2002 sequence occurred in a previously aseismic portion of the slab. Distance is measured along the northwest–southeast ($N80^\circ W$ to $S80^\circ E$).

that occurred eight days earlier at a distance of about 36 km. Eight of the 11 aftershocks located more than 20 km from the main event centroid are consistent with regions of Coulomb stress increase. Similarly, the reported zone of aftershock expansion for the 1994 Tonga deep sequence¹⁵ corresponds to a region where stress was enhanced by the main rupture.

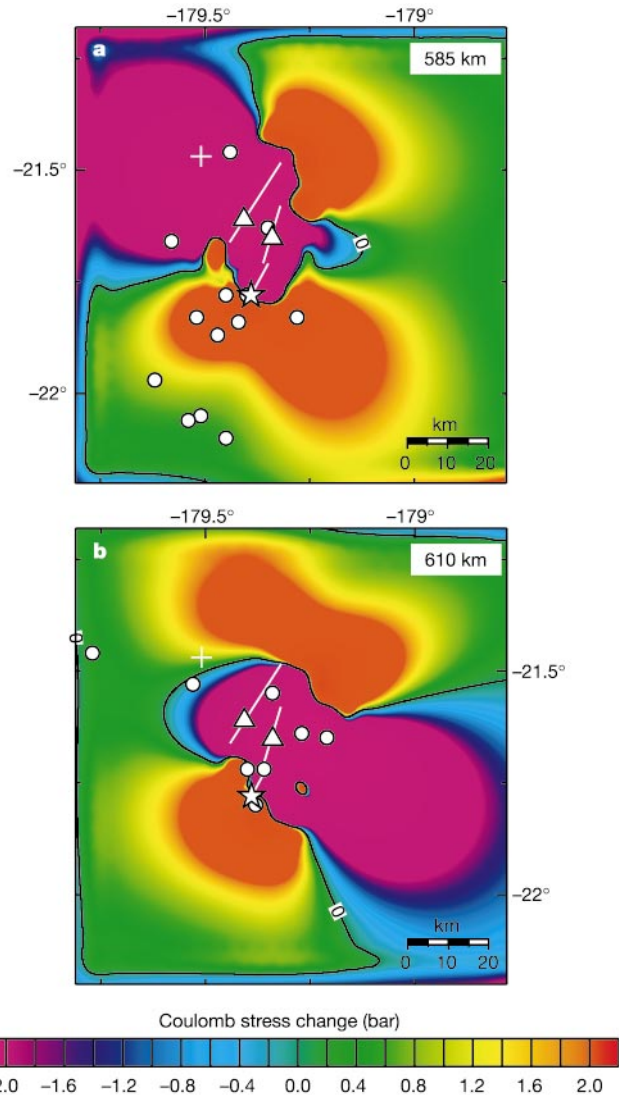


Figure 3 Map showing the location of the first sequence in the 2002 Tonga deep earthquakes, and the calculated static Coulomb stress changes at 585 km (**a**) and 610 km depth (**b**) associated with the mainshock (see Methods). For the short timescale corresponding to the aftershock sequence, the deformation in the relatively rigid subducting plate can be approximated by elastic responses to the fault slip. We applied an approach commonly used for shallow events^{1,32} to compute the co-seismic static stress changes. Because the near-vertical plane ruptured during the first mainshock, the Coulomb stress changes shown have been resolved on a fault whose orientation and slip direction correspond to the average strike, dip and rake of the near-vertical nodal plane of the aftershock focal mechanisms (Fig. 1a). Black lines, zero contours. Star, the relocated mainshock epicentre; triangles, its second and third subevents, as inferred from the body-wave inversion (Supplementary Table 3). Cross, the foreshock; and circles, the aftershocks. Aftershocks plotted in **a** and **b** are those with focal depth ≤ 600 km and > 600 km, respectively. White lines represent the three subfaults in accordance with the best-fitting rupture model of the main event. The fault lengths have been estimated from subevent duration (see Supplementary Fig. 5b) for the best-fitting rupture velocity of 3 km s^{-1} . Note aftershocks south of the mainshock initiation point in **a**. These events are located within and around an area where the Coulomb stress increased by about 2–30 bar.

Calculations of static Coulomb stress changes induced by the first main event on the second mainshock fault show extremely small changes, because the second 2002 Tonga sequence occurred about 300 km (approximately ten source dimensions) away. This can be anticipated from the rapid decay with distance (approximately as $1/r^3$ to $1/r^2$) for static stress changes¹⁶, and indicates that these could not be the triggering agents for the second sequence.

Dynamic strains carried by seismic waves can reduce the frictional strength of faults, promoting their failure and inducing earthquakes¹⁷. Because dynamic strains may occur far from the source, dynamic triggering has been put forward to explain failure at remote distance from an initial earthquake^{4,8}. The northward rupture propagation for the first main event in the 2002 Tonga deep earthquakes is expected to focus seismic waves toward the north, resulting in larger dynamic strains that may favour earthquake triggering in that direction^{16,18}. The location of the triggered sequence rather southwest of the initial earthquake, in an aseismic area, indicates that pre-existing conditions, such as prevailing high stresses, may have been particularly favourable to rupture in that region. The fault must have been in a critical state (close to failure) prior to the arrival of inducing strain pulses.

We have surveyed catalogues of deep earthquakes and include prominent examples of aftershocks triggered at substantial distances in Table 1. Deep earthquakes are commonly triggered in the initial three hours at distances of 50–300 km, well outside the faulting and aftershock zone of the causative event. Static Coulomb stress changes at these distances are small, suggesting that most of these events are dynamically triggered. In most cases, the triggered earthquakes occur in largely aseismic regions, either below or displaced laterally from the active seismic zone. A prominent example is the 26 May 1986 Tonga sequence, in which a M_w 7.1 event occurred in an aseismic region 25 minutes after a M_w 6.8 slab earthquake at a distance of 257 km (Fig. 4). A smaller earthquake occurred further west an additional 42 minutes later. The triggered (outboard) earthquakes are located in a remnant slab that lies

subhorizontally above the active, steeply dipping Tonga deep zone^{19,20}. Aftershock activity continued in the outboard region for 14 days, after which an M_w 7.0 earthquake occurred near the initial event. Assuming a typical rupture length for deep earthquakes with this magnitude²¹ implies that the triggering process operated at a distance of 15–20 fault lengths. The longer duration for the triggered activity suggests that transient processes were not the sole mode of triggering involved. Static loads induced by the triggered M_w 7.1 event may have promoted the later aftershocks in its vicinity.

The occurrence of triggered events in aseismic portions of the slab suggests that these areas may be near criticality. Rupture initiation may be difficult in these regions^{12,22}, such that events are readily triggered by processes following earthquakes in adjacent seismically active zones. These regions can also sustain seismic rupture if an event initiated in the seismically active zone propagates into the surrounding material²². The spatio-temporal separations between the triggering and triggered earthquakes do not define a consistent stress pulse velocity. Instead, the effective propagation velocity of the triggering mechanism varies greatly from a few metres per second up to more than 2 km s^{-1} (Table 1). Because deep earthquakes do not generate surface waves, only body waves are likely to be carriers of dynamic strains. The triggered earthquakes occurred tens of seconds to a few hours after the passage of P and S waves from the causative event, indicating that a short-term delay mechanism may be involved, as observed for shallow earthquakes^{4,6}.

Phenomena such as fluid diffusion, viscoelastic relaxation and fault frictional properties are believed to control delayed triggering for shallow earthquakes^{6,8,23–25}. The first two mechanisms involve longer characteristic times that are inconsistent with the time delays (tens of seconds to a few hours) observed in this study. For shallow faults, rate- and state-dependent friction laws could possibly explain short time delays²⁶. Although frictional laws may not apply directly to deep earthquake failure, other mechanisms with short delay times may be operating. For example, the development of adiabatic

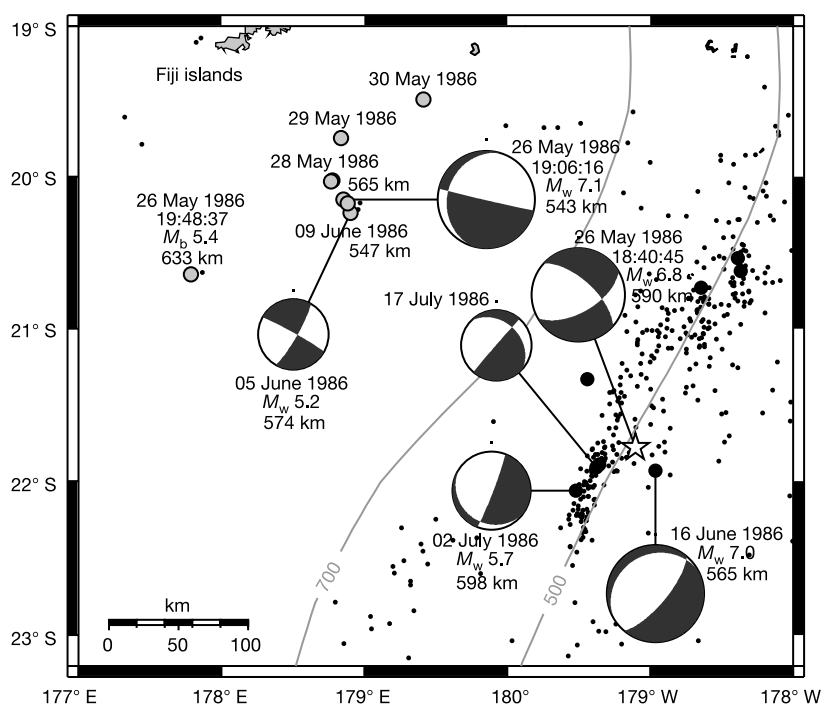


Figure 4 The 1986 Tonga sequence. Star, the initial event. Slab earthquakes are denoted by black circles. Triggered outboard events are represented by grey circles. Focal mechanisms are from the Harvard Centroid Moment Tensor (CMT) catalogue³⁵, with the projection size proportional to the seismic moment. Black and white areas are

compressional and tensional, respectively. The origin time, the magnitude and the depth of the earthquakes are indicated. Dots, the background seismicity³³. Grey lines, contours of deep seismicity³⁴ with the numbers indicating the depth in km to the seismogenic zone.

plastic instabilities^{27–29} may incorporate a time dependence that may serve as a nonlinear temperature-dependent delay factor in dynamic triggering. We propose that deep events can be remotely triggered by transient effects incorporating such nonlinear short-term delay mechanisms in regions where high stress may predominate, but where earthquakes have difficulty nucleating without external influences. □

Methods

Aftershock location

We combined teleseismic arrival times from the Preliminary Determination of Epicentres (PDE) with regional data from the SPANET network (Fig. 1) to constrain the location of the 2002 sequences. P, pP, sP and regional S arrival times were inverted using a hypocentroidal decomposition method that minimizes the effect of velocity heterogeneities along the ray paths³⁰. Two stations of the Global Seismic Network (GSN), AFI and MSVF, and stations of the SPANET network, located at distances of about 5–13° (Fig. 1), recorded upgoing P and/or S phases that allowed better depth constraints.

Body-waveform inversion

Using global broadband data, we inverted for the source parameters of both mainshocks^{21,31}. Global data were combined with recordings of the SPANET network to infer the focal mechanism of the larger ($M_b > 4.6$) aftershocks with a grid search method that fits P and SH waveforms simultaneously.

Coulomb failure stress

The Coulomb failure stress change is expressed as $\Delta\sigma_f = \Delta\tau + \mu' \Delta\sigma$, where $\Delta\tau$ is the shear stress change on a fault (positive in the direction of fault slip), $\Delta\sigma$ is the normal stress change (positive for extension), and μ' is the apparent coefficient of friction, which includes the effects of pore pressure change. Failure is promoted if $\Delta\sigma_f$ is positive and inhibited if negative^{1,32}. The Coulomb stress changes are computed in an elastic half-space using the finite-fault source model for the initial event which consists of three subfaults (Supplementary Table 3). For each subfault, a uniform slip distribution, consistent with the estimated average dislocation of 2 m, is used. Aftershocks located in the depth range between 579 km and 623 km imply widths of about 40 km for the near vertical subfaults. In the calculations, we assumed a shear modulus of 11.6×10^{11} dyn cm⁻², appropriate for a depth of 571 km, and a Poisson's ratio of 0.24 for olivine, the dominant mineral in mantle. For the apparent coefficient of friction μ' , we employed values ranging from 0.6 to 0.9. The general stress distribution was similar for all these values. The final models were computed for $\mu' = 0.7$.

Received 18 April; accepted 14 July 2003; doi:10.1038/nature01903.

- King, G. C. P., Stein, R. S. & Lin, J. Static stress changes and the triggering of earthquakes. *Bull. Seismol. Soc. Am.* **84**, 935–953 (1994).
- Harris, R. A. Introduction to special section: stress triggers, stress shadows, and implications for seismic hazard. *J. Geophys. Res.* **103**, 24347–24358 (1998).
- Stein, R. S. The role of stress transfer in earthquake occurrence. *Nature* **402**, 605–609 (1999).
- Hill, D. P. *et al.* Seismicity remotely triggered by the magnitude 7.3 Landers, California, earthquake. *Science* **260**, 1617–1623 (1993).
- Harris, R. A. & Day, S. M. Dynamics of faults interaction: parallel strike-slip faults. *J. Geophys. Res.* **98**, 4461–4472 (1993).
- Belardinelli, M. E., Cocco, M., Coutant, O. & Cotton, F. Redistribution of dynamic stress during coseismic ruptures: evidence for fault interaction and earthquake triggering. *J. Geophys. Res.* **104**, 14925–14945 (1999).
- Kilb, D., Gombert, J. & Bodin, P. Triggering of earthquake aftershocks by dynamic stresses. *Nature* **408**, 570–574 (2000).
- Gombert, J., Blanpied, M. L. & Beeler, N. M. Transient triggering of near and distant earthquakes. *Bull. Seismol. Soc. Am.* **87**, 295–309 (1997).
- Frohlich, C. The nature of deep-focus earthquakes. *Annu. Rev. Earth Planet. Sci.* **17**, 227–254 (1989).
- Green, H. W. & Houston, H. The mechanics of deep earthquakes. *Annu. Rev. Earth Planet. Sci.* **23**, 169–213 (1995).
- Kirby, S. H., Stein, S., Okal, E. A. & Rubie, D. C. Metastable mantle phase transformations and deep earthquake in subducting oceanic lithosphere. *Rev. Geophys.* **34**, 261–306 (1996).
- Wiens, D. A. Seismological constraints on the mechanism of deep earthquakes: temperature dependence of deep earthquake source properties. *Phys. Earth. Planet. Inter.* **127**, 145–163 (2001).
- Tibi, R., Bock, G. & Wiens, D. A. Source characteristics of large deep earthquakes: constraint on the faulting mechanism at great depths. *J. Geophys. Res.* **108**, doi:10.1029/2002JB001948 (2003).
- Jaeger, J. C. & Cook, N. G. W. *Fundamentals of Rock Mechanics* (Chapman and Hall, London, 1979).
- Wiens, D. W. & McGuire, J. J. Aftershocks of the March 9, 1994, Tonga earthquake: the strongest known deep aftershock sequence. *J. Geophys. Res.* **105**, 19067–19083 (2000).
- Cotton, F. & Coutant, O. Dynamic stress variations due to shear faults in a plane-layered medium. *Geophys. J. Int.* **128**, 676–688 (1997).
- Kisslinger, C. A review of theories of mechanisms of induced seismicity. *Eng. Geol.* **10**, 85–98 (1976).
- Gombert, J., Reasenber, P. A., Bodin, P. & Harris, R. A. Earthquake triggering by seismic waves following the Landers and Hector Mine earthquakes. *Nature* **411**, 462–466 (2001).
- Van der Hilst, R. D. Complex morphology of subducted lithosphere in the mantle beneath the Tonga trench. *Nature* **374**, 154–157 (1995).
- Chen, W.-P. & Brudzinski, M. R. Evidence for a large-scale remnant of subducted lithosphere beneath Fiji. *Science* **292**, 2475–2479 (2001).
- Tibi, R. *Untersuchungen der Ursache von Tiefherdbeben mit Hilfe von Breitbandseismogrammen. Scientific Technical Report STR00/08* (GeoForschungsZentrum Potsdam, Potsdam, 2000).

- McGuire, J. J., Wiens, D. A., Shore, P. J. & Bevis, M. G. The March 9, 1994 (M_w 7.6), deep Tonga earthquake: rupture outside the seismically active slab. *J. Geophys. Res.* **102**, 15163–15182 (1997).
- Hudnut, K. W., Seeber, L. & Pacheco, J. Cross-fault triggering in the November 1987 Superstition Hills earthquake sequence, Southern California. *Geophys. Res. Lett.* **16**, 199–202 (1989).
- Freed, A. M. & Lin, J. Delayed triggering of the 1999 Hector Mine earthquake by viscoelastic stress transfer. *Nature* **411**, 180–183 (2001).
- Pollitz, F. F. & Sacks, I. S. The 1995 Kobe, Japan, earthquake: a long-delayed aftershock of the offshore 1944 Tonankai and 1946 Nankaido earthquakes. *Bull. Seismol. Soc. Am.* **87**, 1–7 (1997).
- King, G. C. P. & Cocco, M. Fault interaction by elastic stress changes: new clues from earthquake sequences. *Adv. Geophys.* **44**, 1–38 (2000).
- Ogawa, M. Shear instability in a viscoelastic material as the cause of deep focus earthquakes. *J. Geophys. Res.* **92**, 13801–13810 (1987).
- Hobbs, B. E. & Ord, A. Plastic instabilities: implications for the origin of intermediate and deep focus earthquakes. *J. Geophys. Res.* **93**, 10521–10521 (1988).
- Karato, S., Riedel, M. R. & Yuen, D. A. Rheological structure and deformation of subducted slabs in the mantle transition zone: implications for mantle circulation and deep earthquakes. *Phys. Earth. Planet. Inter.* **127**, 83–108 (2001).
- Jordan, T. H. & Sverdrup, K. A. Teleseismic location techniques and their application to earthquake clusters in the South-Central Pacific. *Bull. Seismol. Soc. Am.* **71**, 1105–1130 (1981).
- Nábelek, J. L., *Determination of Earthquake Source Parameters from Inversion of Body Waves*. PhD thesis, MIT (1984).
- Toda, S., Stein, R. S., Reasenber, P. A., Dieterich, J. H. & Yoshida, A. Stress transferred by the 1995 M_w = 6.9 Kobe, Japan, shock: effect on aftershocks and future earthquake probabilities. *J. Geophys. Res.* **103**, 24543–24565 (1998).
- Engdahl, E. R., van der Hilst, R. & Buland, R. Global teleseismic earthquake relocation with improved travel times and procedures for depth determination. *Bull. Seismol. Soc. Am.* **88**, 722–742 (1998).
- Gudmundsson, Ö. & Sambridge, M. A regionalized upper mantle (RUM) seismic model. *J. Geophys. Res.* **103**, 7121–7136 (1998).
- Dziwonski, A. M., Chou, T.-A. & Woodhouse, J. H. Determination of earthquake source parameters from waveform data for studies of global and regional seismicity. *J. Geophys. Res.* **86**, 2825–2852 (1981).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank D. Suetsugu and the operators of the SPANET network for their efforts, R. Stein for making his code available, and J. Conder and N. Kagotho for reading the manuscript. Waveforms data were supplied by the Incorporated Research Institutions for Seismology (IRIS), Geoscope, and GeoForschungsNetz (GEOFON). Arrival time data were obtained from the ISC and PDE (US Geological Survey). This research was supported by the National Science Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.T. (tibi@seismo.wustl.edu).

Xenoturbella is a deuterostome that eats molluscs

Sarah J. Boulrat¹, Claus Nielsen², Anne E. Lockyer³, D. Timothy J. Littlewood³ & Maximilian J. Telford¹

¹University Museum of Zoology, Department of Zoology, Downing Street, Cambridge CB2 3EJ, UK

²Zoological Museum (University of Copenhagen), Universitetsparken 15, DK-2100 Copenhagen, Denmark

³Department of Zoology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK

Xenoturbella bocki, first described in 1949 (ref. 1), is a delicate, ciliated, marine worm with a simple body plan: it lacks a through gut, organized gonads, excretory structures and coelomic cavities. Its nervous system is a diffuse nerve net with no brain. *Xenoturbella*'s affinities have long been obscure and it was initially linked to turbellarian flatworms¹. Subsequent authors considered it variously as related to hemichordates and echinoderms owing to similarities of nerve net and epidermal ultrastructure^{2,3}, to acoelomorph flatworms based on body plan and ciliary ultrastructure^{4–6} (also shared by hemichordates⁷), or as among the most primitive of Bilateria⁸. In 1997 two papers seemed to solve this uncertainty: molecular phylogenetic

analyses⁹ placed *Xenoturbella* within the bivalve molluscs, and eggs and larvae resembling those of bivalves were found within specimens of *Xenoturbella*^{10,11}. This molluscan origin implies that all bivalve characters are lost during a radical metamorphosis into the adult *Xenoturbella*. Here, using data from three genes, we show that the samples in these studies were contaminated by bivalve embryos eaten by *Xenoturbella* and that *Xenoturbella* is in fact a deuterostome related to hemichordates and echinoderms.

We carried out independent polymerase chain reaction (PCR) amplifications using ten different sets of primers for three genes (small subunit ribosomal DNA (SSU) and cytochrome oxidase 1 and 2 (*cox1* and 2)) on two specimens of *Xenoturbella* collected from soft mud at a depth of 100 m near Kristineberg, Sweden. From each experiment we found sequences that were most similar to those from bivalve molluscs; in every case we also found another distinct sequence. Our contention that the molluscan sequences are derived from food ingested by *Xenoturbella* is supported by examination of the sequence evidence. Our molluscan-type *cox1* sequence is 97.2% similar at the nucleotide level (593 out of 610) to a sequence from the mollusc *Nucula tenuis* (found at Kristineberg); for comparison *N. tenuis* and congeneric *N. sulcata* are 75% similar. It seems extremely unlikely that *Xenoturbella*, with no molluscan features, could be this genetically similar to a bivalve mollusc. This result is echoed by analyses of a section of SSU (corresponding to positions 1500–1850 in *Homo sapiens* SSU) where we found two molluscan sequences, one identical to the published *Xenoturbella* sequence, the second identical to *N. sulcata* SSU. The most plausible explanation

for these different sequences is that they derive from different bivalve molluscs that have been ingested by *Xenoturbella* as embryos or larvae. This idea is confirmed by experiments comparing SSU PCR products amplified from an intact animal with those in which we dissected away as much of the gut as possible from our specimen. In PCR experiments where mollusc-type sequences are distinguishable from alternative sequences by molecular mass, the concentration of the mollusc-type PCR product decreases in the sample with gut dissected out compared with the intact animal sample; the relative concentration of the alternative sequence increases when the sample contains less gut (Supplementary Information). We conclude from our experiments that the most likely explanation for the similarity to molluscs of some of our sequences and the sequences previously reported from *Xenoturbella* is that they derive from material in the gut and that the alternative sequences must therefore be from *Xenoturbella*.

Phylogenetic analyses^{12,13} using our complete *Xenoturbella* SSU sequence place it within the deuterostomes (Fig. 1; 99% non-parametric bootstrap support (BP) and 100% posterior probability in a bayesian analysis (MB)). Although rooting the Bilateria using different outgroups results in different positions for the root (see Supplementary Information) we make the assumption that the deuterostomes (with or without *Xenoturbella*) are monophyletic. Within the deuterostomes *Xenoturbella* groups with the Ambulacraria (hemichordates and echinoderms) (92% BP, 100% MB) but is basal to both phyla (excluded from Ambulacraria 54% BP, 94% MB). A likelihood ratio test showed that there was significantly more support for positioning *Xenoturbella* within the deuterostomes than for positioning it at the base of the Bilateria or at the base of the deuterostomes ($P < 0.05$). Likelihood ratio tests did not reject the possibility that *Xenoturbella* is a basal hemichordate ($P < 0.061$) or a basal echinoderm ($P < 0.17$) although other information (see below) argues against these possibilities. In our best tree, the urochordates branched at the base of the deuterostomes. Excluding the urochordates from a monophyletic Chordata does not have strong support (76% BP, 86% MB) and a likelihood ratio test shows that our tree with paraphyletic chordates was not significantly better supported than a tree with monophyletic chordates ($P < 0.158$). The observed paraphyly of chordates seems most likely to be a result of the relatively long branches of the urochordates.

We also have nucleotide sequence from *Xenoturbella* correspond-

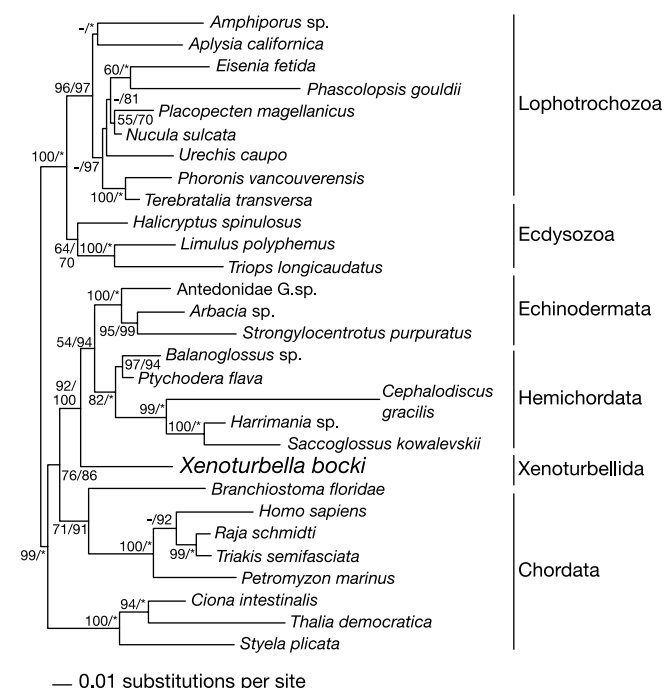


Figure 1 Best ML tree found using SSU sequences. *Xenoturbella* is within the deuterostomes basal to a clade containing the echinoderms and hemichordates. Values are non-parametric bootstrap (first) and MB posterior probabilities (second). MB values of 100 are represented by an asterisk to save space. Maximum parsimony and neighbour-joining analyses placed *Xenoturbella* in an identical position within deuterostomes. The tree has been rooted such that the deuterostomes are monophyletic. Positioning the root on *Xenoturbella* as suggested by rooting with one acoelomorph, *Meara*, would imply paraphyly of the deuterostomes and for this and other reasons is rejected (see Supplementary Information). Forcing *Xenoturbella* to the base of the deuterostomes is significantly less well supported according to likelihood ratio tests ($P < 0.05$). None of the included taxa had a statistically significant divergence from homogeneity of base composition.

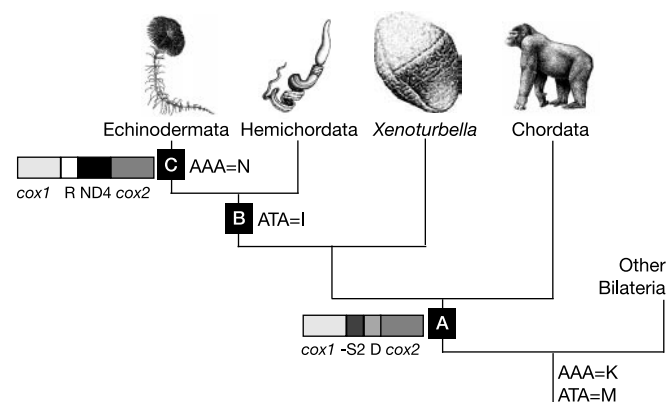


Figure 2 Position of *Xenoturbella* within the deuterostomes as suggested by our analyses of SSU and mitochondrial data. The distribution of synapomorphic molecular character states is indicated by a letter. A, monophyly of deuterostomes including *Xenoturbella* supported by common mitochondrial gene order; B, monophyly of Ambulacraria (hemichordates plus echinoderms) to the exclusion of *Xenoturbella* supported by one genetic code change; C, Monophyly of crown-group echinoderms supported by further genetic code change and gene order change.

ing to most of *cox1*, the 106 amino acids at the amino terminal of *cox2* and the sequence linking these two genes. Although phylogenetic analyses using *cox1* sequences are less reliable at this level of divergence than those of *SSU*, phylogenetic analyses of *cox1* sequences support our results from *SSU*, also placing *Xenoturbella* within the deuterostomes but in this case as the sister group of the hemichordate *Balanoglossus* (see Supplementary Information). These mitochondrial sequences provide us with two additional sources of information for testing the phylogenetic position of *Xenoturbella*—gene order and mitochondrial genetic codes.

Analyses of our mitochondrial sequences¹⁴ show that there are two transfer RNA genes positioned between *Xenoturbella cox1* and *cox2*: *tRNA-S2* (anticodon UGA for serine) adjacent to *cox1* in the reverse orientation, followed by *tRNA-D* (anticodon GUC for aspartic acid) and then *cox2*. Of the metazoans studied so far this gene order is seen only in deuterostomes (cephalochordates, vertebrates and hemichordates; see http://www.jgi.doe.gov/programs/comparative/Mito_top_level.html) providing further evidence that *Xenoturbella* is a deuterostome (Fig. 2). A different gene arrangement is seen in echinoderms demonstrating that *Xenoturbella* is excluded from the crown group of the echinoderms.

An informative mitochondrial genetic code variant is also found in hemichordates and echinoderms¹⁵ (and convergently in rhabdiphoran Platyhelminthes). In both ambulacrarian phyla the codon ATA codes for isoleucine (I) rather than for methionine (M), as is the case in most other metazoans. Examination of *Xenoturbella cox1* and *cox2* following the method in ref. 15 shows that ATA codes for the standard metazoan M (see Supplementary Information). The lack of this hemichordate and echinoderm synapomorphy may suggest that *Xenoturbella* lies outside these two ambulacrarian clades (Fig. 2). There is some further support for this contention, albeit less compelling, from another alteration in mitochondrial genetic code. In echinoderms the codon AAA codes for asparagine (N) rather than lysine (K) as it does in most metazoans, and there has been a concomitant change in the echinoderm lysine tRNA anticodon from TTT to CTT. It has been proposed that the complete absence of the codon AAA from the hemichordate mitochondrial genome is an intermediate step in this reassignment and this is supported by the observation of the same change in the hemichordate lysine tRNA anticodon¹⁶. In contrast, our analyses show that in *Xenoturbella* the codon AAA codes for K as in most other Metazoa (see Supplementary Information).

Xenoturbella cox1 and partial *cox2* genes lack the codons AGA and AGG: the related serine codons AGC and AGT each occur five times. The absence of the AGG codon (if true of the whole *Xenoturbella* mitochondrial genome) is another more tentative link with the deuterostomes; this codon—which codes for serine in most metazoan mitochondria—is not found, or codes for an amino acid other than serine, in all deuterostome taxa apart from echinoderms.

Our results indicate that *Xenoturbella* is not a highly degenerate bivalve mollusc and that the evidence supporting this theory—molluscan eggs, larvae and gene sequences—can be best explained by concluding that *Xenoturbella* eats nudicollid mollusc adults, eggs and pericollymna larvae. Notably, both *N. tenuis* and *N. sulcata* are common in the same habitat as *Xenoturbella*. The published *Xenoturbella SSU* sequence must derive from another of the 109 species of bivalve mollusc found around Kristineberg.

Figure 2 summarizes the relationship of *Xenoturbella* to the other deuterostome taxa suggested by our analyses of *SSU* sequences: in interpreting this tree we make the assumption that the chordates are monophyletic. This scheme of relationships could suggest that the simplicity of *Xenoturbella* represents the primitive form of the deuterostomes and even that this simplicity derives directly from more basal taxa such as the acoelomorphs with which *Xenoturbella* has been linked^{5,6}. It is also possible that *Xenoturbella* is more closely related to one or other of the two ambulacrarian clades, although

this is not seen in our best tree and our mitochondrial code data provide some evidence against this. Whether *Xenoturbella* is basal to the Ambulacraria or sister group to either the hemichordates or the echinoderms, consideration of characters shared by the Ambulacraria and the chordates such as coelomic cavities, complete gut with separate mouth and anus and gill slits, suggest that these characters have been secondarily lost by *Xenoturbella*. This would mean that at least some aspects of its simplicity are derived. Future studies of the embryology of *Xenoturbella* would be expected to reveal further characters commonly found in deuterostomes such as radial cleavage, a ciliated larva and perhaps some degree of left/right asymmetry during development as seen in the ontogeny both of chordates and of hemichordates and echinoderms¹⁷.

Because of the pivotal phylogenetic position of *Xenoturbella* as a close relative of the hemichordates and echinoderms and an out-group to the chordates, future studies of the genetics, morphology and embryology of *Xenoturbella* have the potential to provide great insight into the evolution of the deuterostomes. *Xenoturbella* has previously been placed in the high-ranking taxon Xenoturbellida¹. If future experiments show *Xenoturbella* to be the sister group of the Ambulacraria, as our evidence suggests, then Xenoturbellida would constitute a new phylum of deuterostome. □

Methods

See supplementary Information for specimens, PCR and sequencing primers, and PCR and sequencing methods.

Alignment

SSU sequences were downloaded already aligned according to secondary structure from Ribosomal Database Project II (<http://rdp.cmc.msu.edu/html/>). Our sequence was aligned to these using CLUSTAL X (ref. 18) and the profile alignment option. Alignments were refined by eye using MacClade v. 4.03 (ref. 19).

Exclusion of unreliably aligned positions

To avoid subjectivity in excluding unreliably aligned positions from phylogenetic analyses, the program Gblocks version 0.91b (ref. 20) was used with the standard settings.

Phylogenetic reconstruction

Maximum likelihood (ML) tree estimation used PAUP*4.0b10 (ref. 12) and followed a previously described procedure²¹. PAUP* was also used for maximum parsimony and neighbour joining (NJ) using several methods for calculating the distance matrix.

We also used the MrBayes software¹³ to estimate the best tree using bayesian inference of phylogeny. The gamma parameter with eight rate categories, proportion of invariant sites, GTR matrix and nucleotide frequencies were all estimated. Four chains were run with 1,000,000 generations. Every 100th tree was stored and their likelihoods plateaued after approximately 25% of the run. The last 100 trees (final 10%) were used to make a consensus tree in which the frequency of occurrence of each clade indicates the support for this clade.

Non-parametric bootstraps

NJML non-parametric bootstrapping (NJMLBP) was used to gauge support for relationships. For 1,000 bootstrapped data sets, an NJ tree was calculated from ML distances calculated using the gamma parameter (α) proportion of invariant sites (pinv) parameters fixed as calculated for the original data set based on the ML tree. Base frequencies and rate matrix values were estimated for each bootstrap replicate.

Alternative hypothesis testing with likelihood ratio tests

The Shimodaira–Hasegawa test implemented in PAUP* was used to test whether alternative topologies were significantly less well supported by the data than those found in our ML tree. The RELL approximation method was used with 1,000 bootstrap replicates.

Received 28 February; accepted 9 June 2003; doi:10.1038/nature01851.

- Westblad, E. *Xenoturbella bocki* n.g., n.sp., a peculiar, primitive turbellarian type. *Ark. Zool.* **1**, 3–29 (1949).
- Reisinger, E. Was ist *Xenoturbella*? *Z. Wiss. Zool.* **164**, 188–198 (1960).
- Pedersen, K. J. & Pedersen, L. R. Fine-structural observations on the extracellular-matrix (ecm) of *Xenoturbella bocki* Westblad, 1949. *Acta Zool.* **67**, 103–113 (1986).
- Franzen, A. & Afzelius, B. A. The ciliated epidermis of *Xenoturbella bocki* (Platyhelminthes, Xenoturbellida) with some phylogenetic considerations. *Zool. Scripta* **16**, 9–17 (1987).
- Lundin, K. The epidermal ciliary rootlets of *Xenoturbella bocki* (Xenoturbellida) revisited: new support for a possible kinship with the Acoelomorpha (Platyhelminthes). *Zool. Scripta* **27**, 263–270 (1998).
- Lundin, K. Degenerating epidermal cells in *Xenoturbella bocki* (phylum uncertain), Nemertodermatida and Acoela (Platyhelminthes). *Belg. J. Zool.* **131**, 153–157 (2001).
- Pardos, F. Fine structure and function of pharynx cilia in *Glossobalanus minutus* Kowalewsky (Enteropneusta). *Acta Zool.* **69**, 1–12 (1988).
- Ehlers, U. & Sopott-Ehlers, B. Ultrastructure of the subepidermal musculature of *Xenoturbella bocki*, the adelphotaxon of the Bilateria. *Zoomorphology* **117**, 71–79 (1997).

9. Norén, M. & Jondelius, U. *Xenoturbella's* molluscan relatives... *Nature* **390**, 31–32 (1997).
10. Israelsson, O. ...and molluscan embryogenesis. *Nature* **390**, 32 (1997).
11. Israelsson, O. New light on the enigmatic *Xenoturbella* (phylum uncertain): ontogeny and phylogeny. *Proc. R. Soc. Lond. B* **266**, 835–841 (1999).
12. Swofford, D. L. *Phylogenetic Analysis Using Parsimony (* and other methods) version 4* (Sinauer Associates, Sunderland, Massachusetts, 1998).
13. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* **17**, 754–755 (2001).
14. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964 (1997).
15. Telford, M. J., Herniou, E. A., Russell, R. B. & Littlewood, D. T. J. Changes in mitochondrial genetic codes as phylogenetic characters: two examples from the flatworms. *Proc. Natl Acad. Sci. USA* **97**, 11359–11364 (2000).
16. Castresana, J., Feldmaier-Fuchs, G. & Pääbo, S. Codon reassignment and amino acid composition in hemichordate mitochondria. *Proc. Natl Acad. Sci. USA* **95**, 3703–3707 (1998).
17. Gee, H. *Before the Backbone. Views on the Origin of Vertebrates* (Chapman and Hall, London, 1996).
18. Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G. & Gibson, T. J. Multiple sequence alignment with Clustal X. *Trends Biol. Sci.* **23**, 403–405 (1998).
19. Maddison, D. R., Maddison, W. P. *MacClade 4: Analysis of Phylogeny and Character Evolution* (Sinauer Associates, Sunderland, Massachusetts, 2000).
20. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).
21. Telford, M. J., Lockyer, A. E., Cartwright-Finch, C. & Littlewood, D. T. J. Combined large and small subunit ribosomal RNA phylogenies support a basal position of the acelomorph flatworms. *Proc. R. Soc. Lond. B* **270**, 1077–1083 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank M. Akam and R. Jenner for comments on the manuscript, I. Ruiz Trillo for sharing unpublished results, and the scientists of Kristineberg Marine Station for help in sample collection. We are grateful for support from the Wellcome Trust to M.J.T. and to D.T.J.L.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.J.T. (mt281@cam.ac.uk). Newly determined sequences have been submitted to GenBank under accession numbers AY291292–AY291293.

Speciation by host switch in brood parasitic indigobirds

Michael D. Sorenson^{1,2}, Kristina M. Sefc¹ & Robert B. Payne^{2,3}

¹Department of Biology, Boston University, Boston, Massachusetts 02215, USA

²Museum of Zoology and ³Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan 48109, USA

A growing body of empirical and theoretical work supports the plausibility of sympatric speciation^{1–3}, but there remain few examples in which all the essential components of the process are well understood. The African indigobirds *Vidua* spp. are host-specific brood parasites. Indigobird nestlings are reared along with host young, and mimic the mouth markings of their respective hosts^{4–6}. As adults, male indigobirds mimic host song^{4–7}, whereas females use these songs to choose both their mates and the nests they parasitize⁸. These behavioural mechanisms promote the cohesion of indigobird populations associated with a given host species, and provide a mechanism for reproductive isolation after a new host is colonized. Here we show that all indigobird species are similar genetically, but are significantly differentiated in both mitochondrial haplotype and nuclear allele frequencies. These data support a model of recent sympatric speciation. In contrast to the cuckoo *Cuculus canorus*, in which only female lineages are faithful to specific hosts^{9,10}, host switches have led to speciation in indigobirds because both males and females imprint on their hosts^{8,11}.

The high degree of host specificity in indigobirds led previously to the suggestion that host–parasite associations in African finches were the product of a long history of co-speciation⁴. This model

accounted for the remarkable mimicry of host mouth markings by the young parasites without requiring specialist parasites to have colonized hosts with different mouth markings. Genetic studies, however, indicate that indigobird species have a much more recent origin than their hosts^{12,13}. Indeed, the lack of differentiation among indigobirds in mitochondrial DNA (mtDNA) restriction-fragment length polymorphism (RFLP) markers¹² is somewhat difficult to reconcile with their distinct behaviour¹⁴ and morphology (Fig. 1). Behavioural imprinting in both males (song mimicry)¹¹ and females (mate choice, host choice)⁸ suggests a mechanism for rapid sympatric speciation: indigobirds reared by a novel host species acquire the songs of that host and mate assortatively, resulting in immediate reproductive isolation after a new host is colonized. If this model is correct, the genetic similarity of indigobirds may be attributed to their recent origin from a common ancestor, but evidence of current reproductive isolation also is predicted. We tested this by comparing mitochondrial haplotype and nuclear microsatellite allele frequencies among seven indigobird species in West Africa (samples from Cameroon and Nigeria) and four species in southern Africa (samples from Zimbabwe, Zambia, Malawi and South Africa).

Figure 2 shows unrooted mtDNA haplotype trees for indigobirds. Species within each region share a set of closely related haplotypes, with overall diversity similar to that typically found within a single avian species. For example, a maximum divergence of 2.1% between



Figure 1 Examples of morphological variation between indigobird species. Nestling mouth markings in *V. camerunensis* (a) and *V. chalybeata* (b) mimic the young of their firefinch hosts, *L. rara* and *L. senegala*, respectively. Dark wing and plumage in *V. chalybeata* from West Africa (c). Pale wing and green plumage in *V. raricola* (d). White bill and blue plumage in *V. camerunensis* (e). Red bill and orange feet in *V. chalybeata* from southern Africa (f). See ref. 30 for a complete description of morphological differences between indigobird species.

western and southern indigobird haplotypes is less than that observed within two of their broadly distributed hosts, red-billed firefinch *Lagonosticta senegala* (2.5%) and African firefinch *L. rubricata* (3.9%) (see also Fig. 3). Nonetheless, haplotype frequencies differ significantly between indigobird species, particularly in West Africa (West Africa: $\Phi_{ST} = 0.41$, $P < 0.0001$; southern Africa: $\Phi_{ST} = 0.043$, $P = 0.0014$), providing strong evidence of current reproductive isolation¹⁵. Most pairwise comparisons between individual species were also significant (see Supplementary Information).

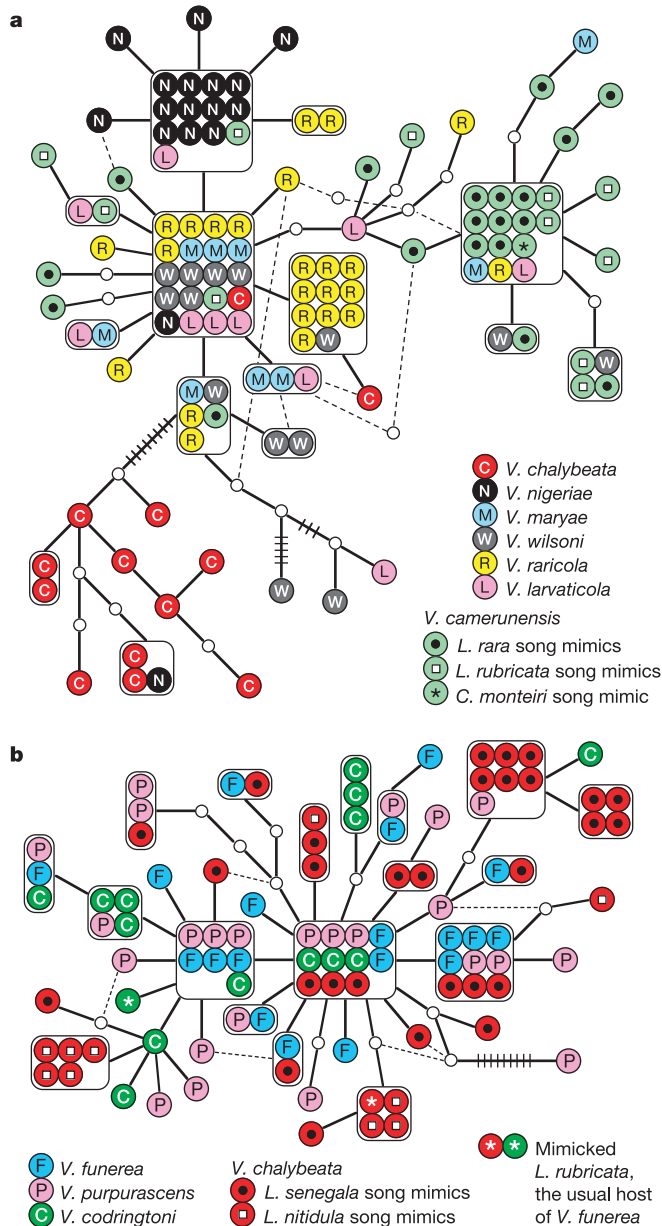


Figure 2 mtDNA haplotype trees for indigobirds. Unrooted phylogenetic trees are shown for 118 indigobirds representing seven morphologically distinct species from Cameroon and Nigeria (**a**), and 98 indigobirds representing four species from southern Africa (**b**). Each circle represents an individual bird. Individuals within a box share the same haplotype. Each line segment represents a single nucleotide substitution, except for branches with cross-marks, which indicate multiple steps. Dotted lines show alternative connections present in one or more of the 3,264 and 1,288 equally parsimonious trees for the two data sets, respectively. One haplotype was found in both regions: the *V. purpurascens* haplotype in the lower right of **b** is identical to the common *V. camerunensis* haplotype in **a**.

In each region, the most common haplotype is shared by all species, and many other haplotypes are derived from it by one or two mutations (Fig. 2). This suggests the retention of a common ancestral haplotype and is consistent with expectations for recent vicariant speciation¹⁵. In indigobirds, however, we suggest that speciation follows the colonization of new host species. If indigobird species retain multiple ancestral lineages, then multiple females must have colonized each new host. An indigobird population along the Zambezi river provides a potential example of this: some village indigobirds *V. chalybeata* in this region are associated with brown firefinch *L. nitidula* rather than their usual host, red-billed firefinch *L. senegala*¹⁶. Indigobirds associated with brown firefinch have four unrelated mtDNA haplotypes, a limited sub-sample of those present in southern birds (Fig. 2b).

Another test of genetic differentiation that emphasizes the colonization process is to treat 'host' as a character and determine the minimum number of changes (that is, host switches) required to explain the current distribution of mtDNA haplotypes. A minimum of 18 host switches is required in southern Africa, whereas at least 22 host switches are required in West Africa, significantly fewer than under a null hypothesis of no genetic structure (southern Africa: $P < 0.006$; West Africa: $P \ll 0.001$), but larger than the minimum possible value of one switch per host.

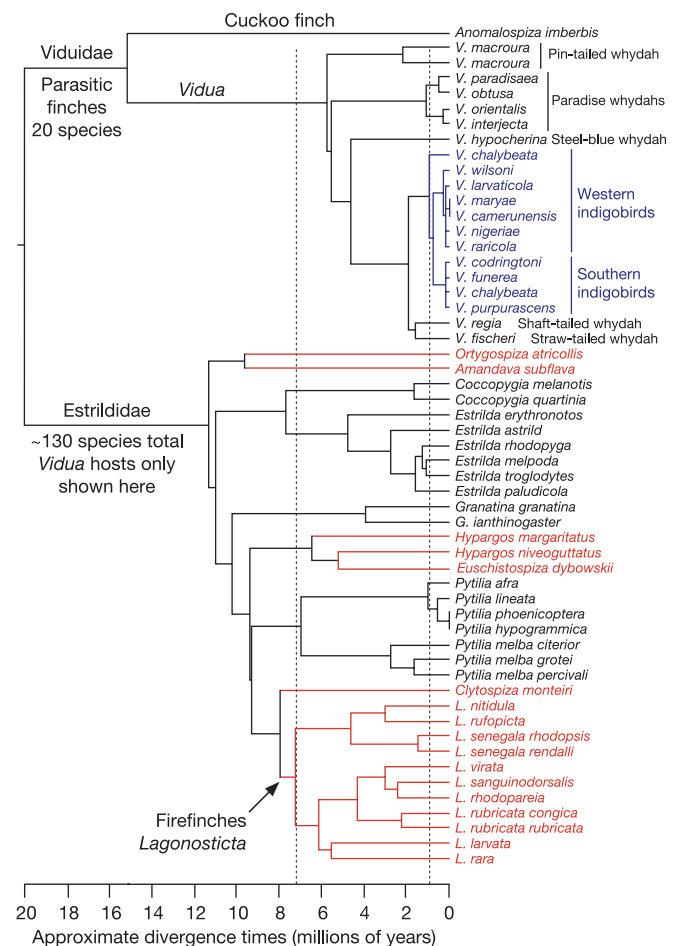


Figure 3 mtDNA phylogeny of brood parasitic finches and their estrildid finch host species. (The cuckoo finch is a parasite of several more distantly related warblers.) Indigobirds are shown in blue; firefinches and other indigobird hosts are shown in red. Other estrildids shown are hosts of the various whydahs. Dotted lines indicate the most recent mtDNA ancestor for indigobirds and firefinches, respectively. Absolute values of divergence times should be viewed as rough approximations at best, but relative times are directly comparable between host and parasitic lineages.

The same behavioural mechanisms involved in host colonization and speciation might also be responsible for occasional hybridization and introgression between established indigobird species. If a female parasitizes a host already associated with a different indigobird species, any surviving offspring will imprint on the alternative host and mate with individuals of the other parasitic species, resulting in hybridization in the second generation. A clear example of relatively recent introgression is the quailfinch indigobird *V. nigeriae* with a haplotype typical of *V. chalybeata* (Fig. 2a, bottom). This individual is the descendant of a *V. chalybeata* female that laid in a quailfinch *Ortygospiza atricapilla* nest, the usual host of *V. nigeriae*. This egg-laying 'mistake' must have occurred at least two and perhaps many more generations in the past, because the individual in question was morphologically *V. nigeriae* and not *V. chalybeata*.

In southern Africa, the mimicry songs of 490 male indigobirds were consistent with their morphology, whereas four males mimicked a species other than the usual host¹⁴. This provides a direct estimate of the proportion of young indigobirds that result from 'misled' eggs (~0.8%) and the potential rate of hybridization among established indigobird species. This is an impressively small number from a behavioural point of view, reflecting the importance of host imprinting in determining the egg-laying behaviour of female indigobirds, but is large by population genetic standards, where the diversifying effects of genetic drift may be counteracted by as few as one migrant per generation ($Nm = 1$)¹⁷. The above model of hybridization predicts no sex bias in 'gene flow' between species (because misled eggs produce both males and females), and therefore differentiation in both mitochondrial and nuclear DNA. Alternatively, frequent mating between male and female indigobirds of different song types would prevent differentiation in nuclear genetic markers even if females remain faithful to hosts in egg laying.

In contrast to recent results for cuckoo host races⁹, nuclear microsatellite data provide evidence of significant genetic differentiation between indigobird species in each region, consistent with assortative mating of indigobirds reared by a particular host species. Overall R_{ST} values (see Methods) were significantly greater than expected under a null hypothesis of no genetic structure (southern Africa: $R_{ST} = 0.027$, $P < 0.0001$; West Africa: $R_{ST} = 0.034$, $P = 0.0015$). Some but not all pairwise comparisons between individual species also were significant (see Supplementary Information). Smaller values of R_{ST} (nuclear microsatellites) than Φ_{ST} (mtDNA) could be taken as evidence of sex-biased gene flow between species. In our view, this interpretation is not warranted, however, because (1) lower differentiation values are expected for highly variable microsatellite markers^{18,19}, and (2) the indigobird system is clearly not at equilibrium—following recent speciation, lineage sorting should proceed more rapidly for mtDNA than for nuclear loci given their different effective population sizes¹⁵.

An alternative to our conclusion of recent speciation is a model in which indigobirds have had long and continuous associations with their current hosts but remain in a perpetual state of incomplete speciation due to gene flow. By impeding differentiation at neutral loci, ongoing gene flow could cause species to appear more recently diverged than they really are¹. Both microsatellite data and nuclear intron sequences, however, are consistent with mtDNA in suggesting limited differentiation among all indigobirds, and an absence of the divergent lineages that would be expected at some loci if indigobirds had a more ancient history (see Supplementary Information). In addition, a phylogeny of parasitic finches indicates a recent divergence of southern and western indigobirds, and of indigobirds from their sister group, straw-tailed whydah *V. fischeri* plus shaft-tailed whydah *V. regia*, placing potential upper limits on the age of southern indigobird species and all indigobirds, respectively (Fig. 3). Finally, southern indigobirds show lower levels of genetic diversity for both mtDNA (Fig. 2) and microsatellites (see

Supplementary Information), consistent with a recent origin from western ancestors (Fig. 3).

Absolute estimates of speciation times are complicated by a lack of relevant fossil information, which is useful for calibrating rates of mtDNA sequence evolution, and the potential for continuing gene flow between species. Using a rough estimate of 20 million years for the divergence of parasitic and estrildid finches²⁰, the common mitochondrial ancestor of all indigobirds occurred less than a million years ago, whereas the history of their primary hosts, the firefinches *Lagonostica* spp., is an order of magnitude longer (Fig. 3). Species divergence times, however, may be substantially more recent than the deepest divergence among extant mtDNA haplotypes²¹. Given an ancient origin of obligate parasitism in finches (Fig. 3), extant indigobirds may represent only the latest iteration of a dynamic process of host colonization, speciation and extinction in the parasitic lineage²⁰.

A remaining question is the evolution and maintenance of host-specific mouth mimicry in nestling indigobirds. Selection by host parents must be strong enough to generate this mimicry, but not so strong as to make occasional host switching impossible. A potential solution is that host discrimination varies with ambient food supply and/or the previous experience of individual hosts^{22,23}, such that selection on mouth markings varies over time or between host nests. Unfortunately, nothing is known of the genetics of mouth markings in indigobirds.

Indigobirds provide an example in vertebrates of sympatric speciation through host shifts, as has been suggested for phytophagous insects². The behavioural mechanisms responsible for reproductive isolation in indigobirds have been tested experimentally and provide a clear explanation for rapid speciation^{8,11}. Sexual selection in the form of female choice is an important component of this process. By mimicking host song, males advertise their success at having been reared by a particular host, whereas females acquire for their offspring compatible mouth mimicry genes by using song to choose their mates. In contrast to models in which divergent ecological selection precedes the evolution of assortative mating³, behavioural imprinting sets the stage for sympatric speciation in indigobirds and facilitates the response to divergent selection on mouth patterns. Behavioural imprinting might also contribute to infrequent hybridization, but significant morphological and genetic differentiation among indigobird species indicates a strong degree of current reproductive isolation. □

Methods

Indigobird samples

During field work between 1991 and 2000, adult male indigobirds were recorded and then trapped using song playback. Adult males ($n = 190$) were identified to species on the basis of morphology and, secondarily, by host song mimicry. Indigobirds were collected from 30 separate locations (>5 km apart) in southern Africa and 23 locations in Cameroon and Nigeria, such that few closely related individuals are included in our sample (see also below). Along the Zambezi River in western Zambia, some *V. chalybeata* mimicked the songs of brown firefinch *Lagonosticta nitidula* rather than those of the usual host, red-billed firefinch *L. senegalensis*¹⁶. In Cameroon, *V. camerunensis* mimicked either African firefinch *L. rubricata* or black-bellied firefinch *L. rara*. In Nigeria, one male *V. camerunensis* mimicked brown twinstot *Clytospiza monteiri*. Finally, one male *V. codringtoni* and one male *V. chalybeata* sang the songs of *L. rubricata* rather than their usual hosts^{12,14} (see Fig. 2). A smaller number of females ($n = 16$) and juveniles ($n = 10$) were identified to species on the basis of association with a given host species (for example, indigobird nestling observed with host) or association with male indigobirds of known species combined with consistent morphology (including mouth markings in juveniles). Samples for genetic analysis included feathers from birds that were individually marked and released, or muscle tissue from birds prepared as museum specimens. Further information on host-parasite associations and the criteria used to recognize indigobird species²⁴ is provided in Supplementary Information.

Genetic data

For each indigobird, we sequenced a 1,100-base-pair region of mtDNA that comprised most of ND6, tRNA-Glu and the 5' half of the control region, using two overlapping primer pairs¹⁶. We are aware of potential problems caused by nuclear copies of mtDNA and are certain that limited genetic differentiation among indigobird species is not an artefact of this phenomenon. Extracts were from muscle tissue or feathers rather than from

blood, and genetic distances from sequence data were consistent with RFLP analyses of purified mtDNA¹². We also scored length variation at 11 nuclear microsatellite loci using primers developed specifically for indigobirds²⁵.

Analyses

Phylogenetic relationships between mtDNA haplotypes were inferred using maximum parsimony as implemented in PAUP^{*26}. Measures of population differentiation (F_{ST}) were calculated using ARLEQUIN²⁷ and RSTCALC²⁸. To compare mtDNA haplotype frequencies between species, we used Φ_{ST} , an F_{ST} analogue that accounts for genetic distances between haplotypes. For microsatellites, we used R_{ST} , which accounts for differences in microsatellite repeat number under a stepwise mutation model. Standard F -statistics lead to identical conclusions. Significance was assessed using permutation procedures implemented in the respective programs. We excluded from analyses of population structure two individuals that may have been close genetic relatives of another individual in our sample, on the basis of shared location and mtDNA haplotype, and significantly greater microsatellite allele sharing than expected given population-level frequencies.

Significant genetic structure among indigobird species was not an artefact of geographic structure combined with sampling different species in somewhat different areas. Partial Mantel tests controlling for differences in allele frequency between species suggest no relationship between geographic distance and microsatellite allele sharing (West Africa: $r = -0.004$, $P = 0.34$; southern Africa: $r = 0.009$, $P = 0.21$). By contrast, allele sharing within species is greater than between species even when controlling for geographic distance between samples (West Africa: $r = -0.047$, $P < 0.001$; southern Africa: $r = -0.019$, $P = 0.052$; K.M.S. *et al.*, unpublished data).

To estimate the minimum number of host switches needed to explain the distribution of mtDNA haplotypes among indigobird species, the minimum number of steps in the multistate character 'host' was determined for the most parsimonious trees when 'host' was included as an additional character during tree search. To evaluate if the minimum number of host switches was smaller than expected under a null model of no association between host species and indigobird mtDNA haplotype, a null distribution was generated by reassigning individuals to hosts and determining the minimum number of host switches in 1016 replicate analyses.

To put the indigobird radiation in a broader phylogenetic context, we analysed mtDNA sequence data for representative indigobirds, other parasitic finches and their estrildid hosts. The phylogeny presented here (Fig. 3) is based on a maximum-likelihood analysis of 1,563 aligned positions, including the regions noted above plus half of the ND2 gene. Non-host estrildids were pruned from the tree, and branch lengths were estimated in PAUP^{*26} under a GTR + I + Γ model of sequence evolution. We estimated relative divergence times using the Langley–Fitch method, as implemented in the program r8s²⁹, using a single calibration point of 20 million years for the divergence of parasitic and estrildid finches²⁰. Local molecular clocks were specified for *Vidua*, *Anomalospiza* and estrildids, respectively, to account for a faster rate of sequence evolution in parasitic finches²⁰.

Received 14 February; accepted 23 June 2003; doi:10.1038/nature01863.

1. Via, S. Sympatric speciation in animals: the ugly duckling grows up. *Trends Ecol. Evol.* **16**, 381–390 (2001).
2. Berlocher, S. H. & Feder, J. L. Sympatric speciation in phytophagous insects: moving beyond controversy? *Annu. Rev. Entomol.* **47**, 773–815 (2002).
3. Dieckmann, U. & Doebeli, M. On the origin of species by sympatric speciation. *Nature* **400**, 354–357 (1999).
4. Nicolai, J. Der Brutparasitismus der Viduinae als ethologisches Problem. *Z. Tierpsychol.* **21**, 129–204 (1964).
5. Payne, R. B. Behavior, mimetic songs and song dialects, and relationships of the parasitic indigobirds (*Vidua*) of Africa. *Ornithol. Monogr.* **11**, 1–333 (1973).
6. Payne, R. B. Species limits in the indigobirds (*Ploceidae*, *Vidua*) of West Africa: mouth mimicry, song mimicry, and description of new species. *Misc. Publ. Univ. Mich. Mus. Zool.* **162**, 1–96 (1982).
7. Payne, R. B. & Payne, L. L. Song mimicry and species associations of West African indigobirds *Vidua* with quail-finch *Ortygospiza atricollis*, goldbreast *Amandava subflava* and brown twospot *Clytospiza monteiri*. *Ibis (Lond. 1859)* **136**, 291–304 (1994).
8. Payne, R. B., Payne, L. L., Woods, J. L. & Sorenson, M. D. Imprinting and the origin of parasite-host species associations in brood-parasitic indigobirds, *Vidua chalybeata*. *Anim. Behav.* **59**, 69–81 (2000).
9. Gibbs, H. L. *et al.* Genetic evidence for female host-specific races of the common cuckoo. *Nature* **407**, 183–186 (2000).
10. Marchetti, K., Nakamura, H. & Gibbs, H. L. Host-race formation in the common cuckoo. *Science* **282**, 471–472 (1998).
11. Payne, R. B., Payne, L. L. & Woods, J. L. Song learning in brood-parasitic indigobirds *Vidua chalybeata*: song mimicry of the host species. *Anim. Behav.* **55**, 1537–1553 (1998).
12. Klein, N. K. & Payne, R. B. Evolutionary associations of brood parasitic finches (*Vidua*) and their host species: Analyses of mitochondrial DNA restriction sites. *Evolution* **52**, 566–582 (1998).
13. Sorenson, M. D. & Payne, R. B. Molecular genetic perspectives on avian brood parasitism. *Integr. Comp. Biol.* **42**, 388–400 (2002).
14. Payne, R. B., Payne, L. L., Nhlane, M. E. D. & Hustler, K. Species status and distribution of the parasitic indigo-birds *Vidua* in east and southern Africa. *Proc. VIII Pan-Afr. Ornithol. Congr.*, 40–52 (Bujumbura, Burundi, 1993).
15. Avise, J. C. *Phylogeography: The History and Formation of Species* (Harvard Univ. Press, Cambridge, Massachusetts, 2000).
16. Payne, R. B., Hustler, K., Stjernstedt, R., Sefc, K. M. & Sorenson, M. D. Behavioural and genetic evidence of a recent population switch to a novel host species in brood-parasitic indigobirds *Vidua chalybeata*. *Ibis (Lond. 1859)* **144**, 373–383 (2002).
17. Slatkin, M. Gene flow and the geographic structure of natural populations. *Science* **236**, 787–792 (1987).
18. Balloux, F. & Lugon-Moulin, N. The estimation of population differentiation with microsatellite markers. *Mol. Ecol.* **11**, 155–165 (2002).
19. Hedrick, P. W. Perspective: highly variable loci and their interpretation in evolution and conservation. *Evolution* **53**, 313–318 (1999).

20. Sorenson, M. D. & Payne, R. B. A single ancient origin of brood parasitism in African finches: implications for host–parasite coevolution. *Evolution* **55**, 2550–2567 (2001).
21. Edwards, S. V. & Beerli, P. Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution* **54**, 1839–1854 (2000).
22. Lotem, A., Nakamura, H. & Zahavi, A. Constraints on egg discrimination and cuckoo host coevolution. *Anim. Behav.* **49**, 1185–1209 (1995).
23. Payne, R. B., Woods, J. L. & Payne, L. L. Parental care in estrildid finches: experimental tests of a model of *Vidua* brood parasitism. *Anim. Behav.* **62**, 473–483 (2001).
24. de Queiroz, K. in *Endless Forms: Species and Speciation* (eds Howard, D. & Berlocher, S. H.) 57–75 (Oxford Univ. Press, Oxford, 1998).
25. Sefc, K. M., Payne, R. B. & Sorenson, M. D. Characterization of microsatellite loci in village indigobirds *Vidua chalybeata* and cross-species amplification in estrildid and ploceid finches. *Mol. Ecol. Notes* **1**, 252–254 (2001).
26. Swofford, D. L. *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*, Version 4.0b10 (Sinauer Associates, Sunderland, Massachusetts, 2002).
27. Schneider, S. D., Roessli, D. & Excoffier, L. *ARLEQUIN: A Software for Population Genetics Analysis* (Univ. Geneva, Switzerland, 2000).
28. Goodman, S. J. *RST Calc: a collection of computer programs for calculating estimates of genetic differentiation from microsatellite data and determining their significance*. *Mol. Ecol.* **6**, 881–885 (1997).
29. Sanderson, M. J. *r8s, Version 1.60* (Univ. California Davis, California, 2003).
30. Payne, R. B. Field identification of the indigobirds. *Bull. Afr. Bird Club* **3**, 14–25 (1996).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements For comments, logistical support, and assistance in the field and lab, we thank Laura Payne, Chris Balakrishnan, Clive Barlow, Adrian Craig, Nick Davies, Roger Fotoso, Kathy Groschupf, Janet Hinshaw, Mark Hopkins, Kit Hustler, Lacey Knowles, Kevin Njabo, Nedra Klein, David Mindell, and Bob Stjernstedt. Funding was provided by US National Science Foundation Grants to R.B.P. and M.D.S. and an Erwin Schrödinger Fellowship from the Austrian Science Fund for K.M.S.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.D.S. (msoren@bu.edu). DNA sequence data have been deposited in GenBank under accession numbers AF090341, AF407021–AF407036, AF407107–AF407122, AY136611, AY322613–AY322837, AY323530–AY323607 and AY324232–AY324306.

Genetic mechanisms and constraints governing the evolution of correlated traits in drosophilid flies

Nicolas Gompel & Sean B. Carroll

Howard Hughes Medical Institute and Laboratory of Molecular Biology, University of Wisconsin, 1525 Linden Drive, Madison, Wisconsin 53706, USA

Some morphological traits differ greatly between related species, but it is not clear whether diversity evolves through changes in the same genes and whether similar, independent (that is, convergent) changes occur by the same mechanism^{1,2}. Pigmentation in fruitflies presents an attractive opportunity to explore these issues because pigmentation patterns are diverse, similar patterns have arisen in independent clades, and numerous genes governing their formation have been identified^{3–5} in *Drosophila melanogaster*. Here we show that both evolutionary diversification and convergence can be due to evolution at the same locus, by comparing abdominal pigmentation and trichome patterns and the expression of *Bric-à-brac2* (*Bab2*), which regulates both traits in *D. melanogaster*^{3,6}, in 13 species representing the major clades^{7,8} of the subfamily Drosophilinae. Modifications of *Bab2* expression are frequently correlated with diverse pigmentation and trichome patterns that evolved independently in multiple lineages. In a few species, *Bab2* expression is not correlated with changes in pigmentation but is correlated with a conserved pattern of trichomes, indicating that this locus can be circumvented to evolve new patterns when a correlated trait is under different constraints.

In *D. melanogaster*, the abdominal melanic pigmentation pattern comprises a banding pattern (found in most Drosophilinae), in which each abdominal segment bears a dark stripe, with a superimposed dimorphic pattern in which males are darkly pigmented on their terminal segments A5 and A6 (Fig. 1a). This dimorphic pattern is controlled by the products of the *bric-à-brac* (*bab*) locus³, which contains two largely redundant genes^{3,6}, *bric-à-brac1* (*bab1*) and *bric-à-brac2* (*bab2*) derived from a tandem duplication. Both genes are expressed sex-specifically and segment-specifically in

the pupal abdomen during epidermal development³, during which the Bab proteins act as repressors of pigmentation. It has been shown³ that dimorphic pigmentation associated with the posterior male-specific repression of *bab* evolved within the *D. melanogaster* species group (a group of closely related species within the subgenus *Sophophora*⁹; Fig. 2).

Because other species of the *Sophophora* subgenus are monomorphic with regard to pigmentation and *bab* regulation, as well as more distant species outside the *Sophophora*, we had no *a priori* expectation that *bab* had a role in pigmentation divergence in any group outside the *melanogaster* species group. We were therefore surprised to find that Bab2 expression is modulated in a variety of patterns in other members of the Drosophilinae (Fig. 1; see Fig. 2 for phylogeny). A survey of 13 species with conspicuous and diverse melanic pigmentation patterns representing the major clades of this subfamily^{7,8} revealed a spectrum of species-specific modulations of Bab2 expression in the dorsal epidermis of developing pupal abdomens (Fig. 2). Four modes of Bab2 regulation occur in the subfamily that are well correlated with adult patterns of pigmentation, including monomorphic posterior repression (*D. immigrans*, Fig. 1g), elevated expression along the dorsal midline (*D. funebris*, Fig. 1e; *D. immigrans*, Fig. 1h), and repression in territories along the dorsal midline (*D. tripunctata*, Fig. 1k; *D. duncani*, Fig. 1n). The male-specific posterior repression of Bab2 associated with pigmentation found in species of the subgenus *Sophophora* (*D. melanogaster*, Fig. 1b) has evolved repeatedly in the subfamily (*D. funebris*, Fig. 1d; *D. duncani*, Fig. 1m). Bab2 expression is further diversified by the combination of these four modes of regulation. For instance, the posterior repression of Bab2 can be dimorphic or not, and associated or not with modulation in expression along the midline (for example, *D. duncani* combines dimorphic and midline repression of Bab2).

The correlation between Bab2 expression and the repression of melanic pigmentation ranges from a very close correspondence between Bab2 expression and areas of no pigmentation in species such as *D. melanogaster* or *D. tripunctata* (Fig. 1a, b, i, j, k) to a partial correlation in other species in which Bab2 relates to some aspects of the pigmentation, such as *D. phalerata* (dimorphic posterior pigmentation, Fig. 2) but other regulators are necessary to explain the entire pattern (midline repression). In regions where Bab2 expression prefigures the absence of pigmentation, it is not strictly repressed in the prospective pigmented areas and activated elsewhere; rather, pigmentation seems to be repressed where Bab2 is expressed above a threshold (Fig. 1c, e, f, h, i, k, n). The widespread association of Bab2 expression with repression of adult pigmentation suggests that Bab2 functions as a regulator of pigmentation in most species and was expressed in the abdominal epidermis of a common ancestor.

Importantly, however, in 3 of 13 species no correlation exists between Bab2 expression and pigmentation. In *D. santomea* (Figs 2 and 3a), a close relative to *D. melanogaster*, both sexes are devoid of dark pigments. This is a recent loss of the melanic pattern, because its sister species *D. yakuba* is dimorphic¹⁰. One might expect Bab2 to be expressed uniformly throughout the abdomen in *D. santomea* to repress pigmentation, but that is not so. Bab2 is expressed in a dimorphic *melanogaster*-like pattern (Fig. 3a). The loss of pigmentation in *D. santomea* must therefore be due to evolution at other loci. The conservation of Bab2 expression is consistent with a recent genetic association study that ruled out the *bab* locus as contributing to differences in pigmentation between *D. yakuba* and *D. santomea*¹¹.

However, this raises the question of why Bab2 is regulated dimorphically in this species if it is not correlated with pigmentation. We found that another abdominal character, the pattern of trichomes, is perfectly correlated with Bab2 expression in *D. santomea* (Fig. 3a). Trichomes are small non-sensory hairs that cover the cuticle (Fig. 3b, inset). In *D. melanogaster*, *bab* function is

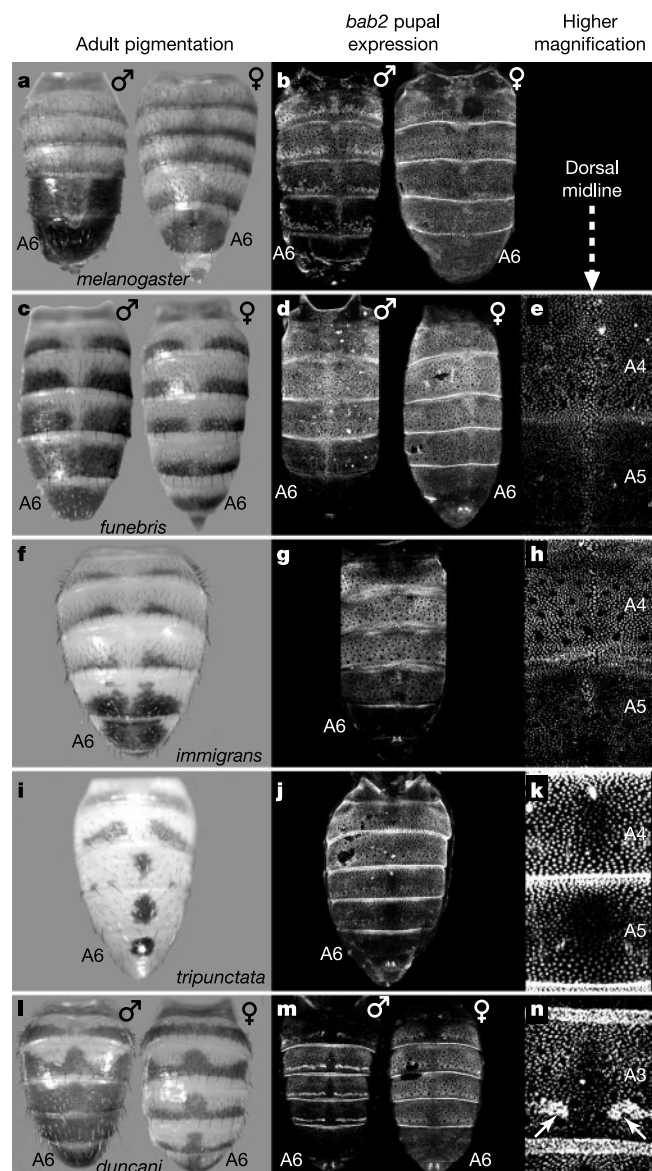


Figure 1 Modulation of Bab2 expression is correlated with diverse abdominal pigmentation patterns. **a, c, f, i, l**, Dorsal view of abdomens of *D. melanogaster* (**a**), *D. funebris* (**c**), *D. immigrans* (**f**), *D. tripunctata* (**i**) and *D. duncani* (**l**) adults. Pigmentation varies between species in the posterior segments, along the dorsal midline and between sexes. **b, d, g, j, m**, Bab2 expression in the developing adult epidermis of each species revealed by immunostaining is correlated with the repression of adult pigmentation. **e, h, k, n**, Higher-magnification views of Bab2 immunostainings along the dorsal midline (indicated by large arrow). Note that in the *D. tripunctata* adult, the spot on A4 is smaller than those on A5 and A6; this is related to the area and level of Bab2 repression. In *D. melanogaster* and *D. funebris*, the Bab2 protein concentration is somewhat lower but still detectable in A5 and A6 in the females, whereas it is faint in A5 and undetectable in A6 in the males. The small arrows in **n** point to a non-epidermal structure, the oenocytes, that express high levels of Bab2.

required for both pigmentation repression and trichome development³. The *melanogaster*-like pattern of Bab2 expression in *D. santomea* suggests that *bab* controls these traits independently and that their evolution can be uncoupled.

A second, illuminating example of a lack of correlation between Bab2 expression and pigmentation is found in *D. serrata*, another close relative of *D. melanogaster*. Pigmentation is dimorphic in some populations of this species¹², but the pattern is opposite to the *D. melanogaster* pattern: the male is pale and the female is darkly pigmented on A6 (Figs 2 and 3b). Despite this pattern reversal

between the sexes, we found a *melanogaster*-like Bab2 pattern in which posterior repression occurs in the male. This pattern is also correlated with the trichome distribution (Fig. 3b). We infer that Bab2 does not repress pigmentation in this species, which allows dark pigmentation of A6 in the female. The expanded posterior pigmentation in *D. serrata* females is a convergent pattern that must have evolved through a different genetic pathway, independent of *bab*, indicating that phenotypic convergence does not necessarily rely on the same genetic mechanism^{1,13}. We suggest that control of the trichome pattern by *bab2* has been maintained in *D. melanogaster*,

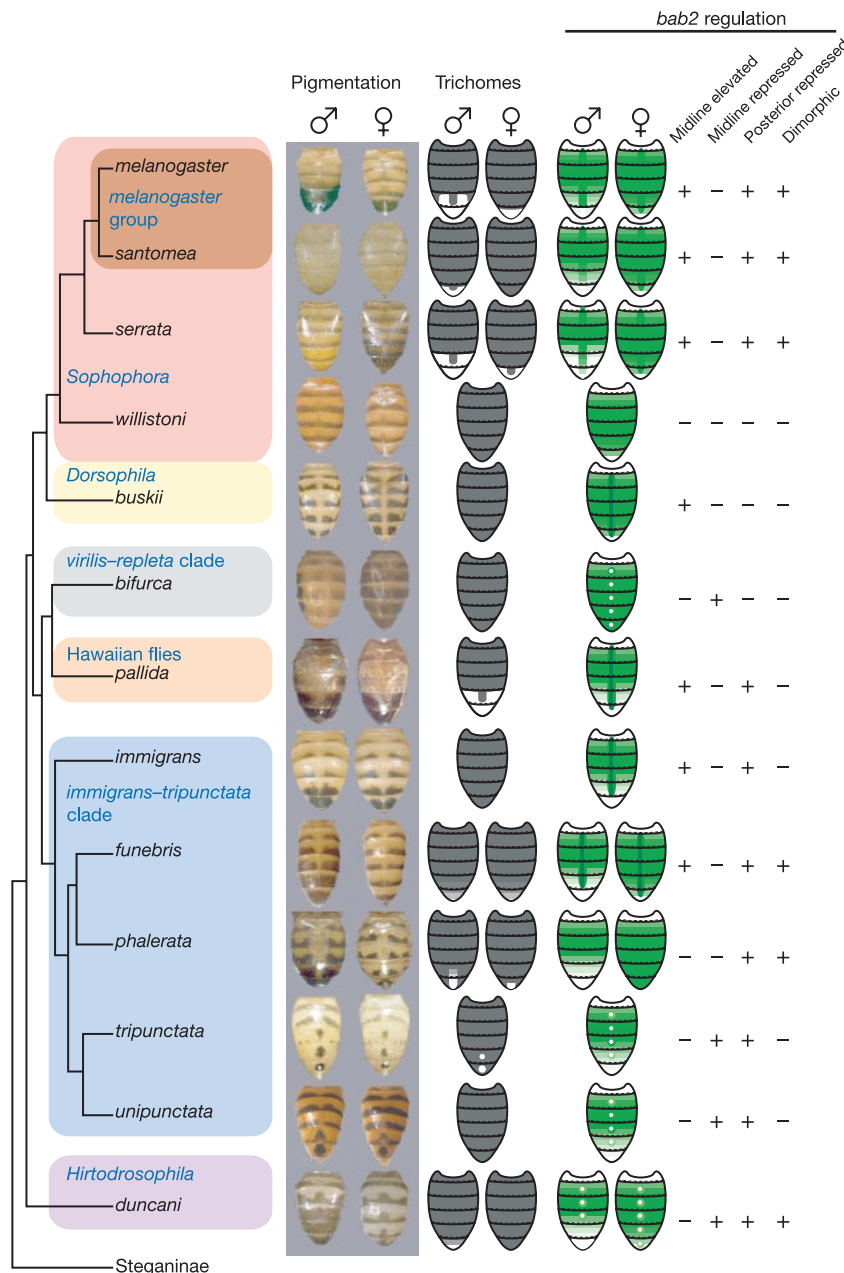


Figure 2 Modulation of Bab2 expression underlies the diversification and evolutionary convergence of cuticular traits throughout the Drosophilinae. Depicted are the phylogeny of the Drosophilinae, the variety of abdominal pigmentation patterns, Bab2 expression in respective species and sexes (green shading), the trichome patterns (grey shading) and the different modes of *bab2* regulation observed in each species (+ or - for presence or absence, respectively). Left, the phylogeny of the Drosophilidae shown here is adapted from ref. 8. There are conflicts in phylogenies of the family^{7,8,24} and none might reflect the actual evolutionary tree. We relied on the two largest phylogenies of the

family, one based on morphology⁷, the other on DNA sequences⁸, to reconstruct (using parsimony assumptions) the four changes observed in *bab2* regulation with MacClade version 4.03 (ref. 25) and found that characters have changed multiple times (13 changes in all for each phylogeny, but with different distributions). Note that trichomes are present on A1 in all species, although Bab2 is consistently repressed in this segment throughout the subfamily, except in some species that express Bab2 along the dorsal midline.

D. santomea and *D. serrata* because their absence from the terminal segments in males has some biological role.

The correlation between Bab2 expression and trichome patterns holds in many of the species we examined (9 of 13 species, for example *Scaptomyza pallida*, Fig. 3c; *D. tripunctata*, Fig. 3d), indicating that in most cases the two traits have evolved in concert, apparently under control of *bab*. However, there are a few exceptions such as *D. immigrans* (Fig. 3e) and *D. unipunctata* (Fig. 2), which are covered by trichomes in regions where Bab2 is strongly down-regulated, indicating that trichome pattern evolution can also be uncoupled from Bab2.

In species in which patterns have become uncoupled from Bab2 (*D. santomea* and *D. serrata*), it is reasonable to infer that each trait might be under selective pressures that are in conflict with respect to Bab2 function (for example, to become darker on A6 while preserving trichome formation on that segment). Alleles at other loci affecting one of the traits might be selected so that trait evolution might circumvent Bab2 function. In natural populations of *D. melanogaster*, variation at the *bab* locus is associated with

about 60% of the variation in female posterior pigmentation¹⁴, indicating that 40% of the variation is contributed by other loci. Such loci could evolve as major regulators of pigmentation if *bab* is constrained by a stable trichome pattern (as in *D. santomea* or *D. serrata*).

Our results indicate that Bab2 regulation has been modified repeatedly, leading to its modulation within the abdominal field, which seems to be the basis of much of the abdominal pattern diversification in this subfamily (Fig. 2). We suggest that changes *in cis* to *bab2*, in regulatory elements responding directly or indirectly to one or more body plan regulators (*AbdominalB* and *doublesex* for segment specificity or sex specificity³, possibly *pannier*¹⁵ or *decapentaplegic*¹⁶ for regulation along the dorsal midline), are responsible for the diversification of Bab2 expression. It is less likely that changes in the spatial deployment of these regulators account for differences in Bab2 patterns between species, because comparative data indicate that these regulators of body pattern are strongly constrained and that their expression or function is usually conserved at the family or order level^{17–19}. Isolation and analysis of *bab2* regulatory elements from different species will be required to determine how *bab* regulation has evolved.

Evolution of Bab2 expression underlies most cases of phenotypic convergence in our survey. Character reconstruction (legend to Fig. 2) shows that all of the regulatory changes in *bab2* have evolved more than once when we consider either of the two major phylogenies of the family^{7,8}. We suggest that convergence was achieved through similar, independent regulatory changes in *bab*. Other animal groups also exhibit melanin variation on one or a few basic themes, such as the spot patterns on the forewings of ladybirds. Alleles controlling many different patterns that occur in the 2-spot ladybird map to a single locus^{20,21}, indicating that different alleles represent regulatory variants and that similar ladybird patterns could arise through a common genetic mechanism. We suggest that the flexible modulation of a transcription factor that regulates colour patterns might be a general mechanism governing the diversity of body-colour patterns in related animal species. □

Methods

Breeding of flies

Flies were bred on cornmeal medium sometimes supplemented with banana and kept at 20 °C. Staging follows ref. 22 for *D. melanogaster*. A similar developmental stage was identified empirically for other species developing at various rates.

Immunocytochemistry

Epidermis from the dorsal abdomen of pupae were dissected (fat body was cleaned off) in PBS, and stained with an anti-Bab2 antibody³ and a fluorescein isothiocyanate-conjugated anti-rat IgG antibody (Jackson ImmunoResearch). Preparations were processed for confocal imaging on an Optiphot microscope (Nikon) equipped with a ×10 dry lens and a Bio-Rad 1024 system. In all species, Bab2 is expressed in the cells that prefigure the adult epidermis and the highest level of expression occurs at a stage equivalent to 38–42 h after puparium formation in *Drosophila melanogaster*²².

Cuticle preparation

Adult cuticles were cleared for 1 h in 10% KOH at 60 °C, washed in water and mounted flat in Hoyer's medium²³, then processed for dark-field imaging with a ×10 dry lens on a Zeiss Axiophot microscope equipped with a Kontron charge-coupled-device camera.

Received 26 March; accepted 14 May 2003; doi:10.1038/nature01787.

1. Wray, G. A. Do convergent developmental mechanisms underlie convergent phenotypes? *Brain Behav. Evol.* **59**, 327–336 (2002).
2. Maynard Smith, J. et al. Developmental constraints and evolution. *Q. Rev. Biol.* **60**, 265–287 (1985).
3. Kopp, A., Duncan, I., Godt, D. & Carroll, S. B. Genetic control and evolution of sexually dimorphic characters in *Drosophila*. *Nature* **408**, 553–559 (2000).
4. Kopp, A. & Duncan, I. Control of cell fate and polarity in the adult abdominal segments of *Drosophila* by optomotor-blind. *Development* **124**, 3715–3726 (1997).
5. Wittkopp, P. J., True, J. R. & Carroll, S. B. Reciprocal functions of the *Drosophila* yellow and ebony proteins in the development and evolution of pigment patterns. *Development* **129**, 1849–1858 (2002).
6. Couderc, J. L. et al. The *bric-à-brac* locus consists of two paralogous genes encoding BTB/POZ domain proteins and acts as a homeotic and morphogenetic regulator of imaginal development in *Drosophila*. *Development* **129**, 2419–2433 (2002).
7. Grimaldi, D. A. A phylogenetic, revised classification of genera in the Drosophilidae (Diptera). *Bull. Am. Mus. Nat. Hist.* **197**, 1–139 (1990).
8. Remsen, J. & O'Grady, P. Phylogeny of Drosophilinae (Diptera: Drosophilidae), with comments on

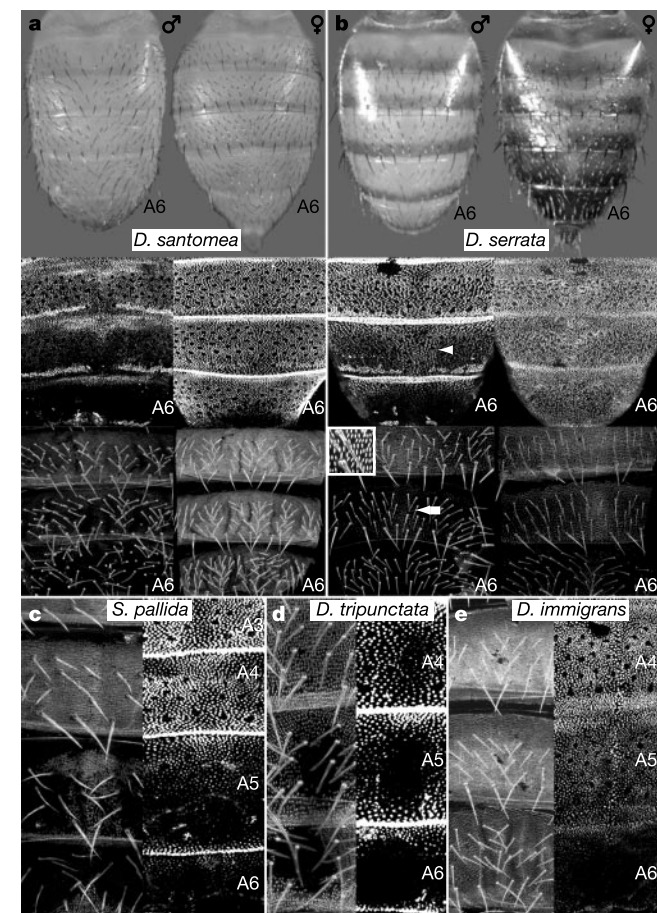


Figure 3 Bab2 expression is correlated with the distribution of cuticular trichomes. **a**, *D. santomea* adults are devoid of melanin pigment (top). Bab2 distribution (middle) is not correlated with repression of pigmentation, but is correlated with the presence of trichomes (bottom). Trichomes are small epidermal productions spread amid the sensory bristles (higher-magnification inset in **b**, bottom). **b**, *D. serrata* adults exhibit dimorphic melanism (top) but with a sexual pattern the reverse of *D. melanogaster*. Bab2 is expressed in a *melanogaster*-like pattern (middle) which is not correlated with pigmentation but prefigures the pattern of trichomes. Trichomes are absent from A6 and part of A5 in the males, but do form along the dorsal midline (arrow) where Bab2 expression is elevated (arrowhead). **c–e**, The correlation between trichomes (left panels) and Bab2 expression (right panels) is general in Drosophilinae as exemplified in *S. pallida* (**c**) and *D. tripunctata* (**d**). Occasionally, however, trichomes appear in areas where Bab2 is strongly repressed, as in the A6 segment in *D. immigrans* (**e**).

- combined analysis and character support. *Mol. Phylogenet. Evol.* **24**, 249–264 (2002).
9. Bock, I. R. & Wheeler, M. R. in *Studies in Genetics* (ed. Wheeler, M. R.) 1–102 (Univ. of Texas, Austin, 1972).
 10. Lachaise, D. *et al.* Evolutionary novelties in islands: *Drosophila santomea*, a new *melanogaster* sister species from Sao Tome. *Proc. R. Soc. Lond. B* **267**, 1487–1495 (2000).
 11. Llopart, A., Elwyn, S., Lachaise, D. & Coyne, J. A. Genetics of a difference in pigmentation between *Drosophila yakuba* and *Drosophila santomea*. *Evolution* **56**, 2262–2277 (2002).
 12. Ayala, F. J. Sibling species of the *Drosophila serrata* group. *Evolution* **19**, 538–545 (1965).
 13. Wittkopp, P. J., Williams, B. L., Selegue, J. E. & Carroll, S. B. *Drosophila* pigmentation evolution: Divergent genotypes underlying convergent phenotypes. *Proc. Natl Acad. Sci. USA* **100**, 1808–1813 (2003).
 14. Kopp, A., Graze, R. M., Xu, S., Carroll, S. B. & Nuzhdin, S. V. Quantitative trait loci responsible for variation in sexually dimorphic traits in *Drosophila melanogaster*. *Genetics* **163**, 771–787 (2003).
 15. Calleja, M. *et al.* Generation of medial and lateral dorsal body domains by the pannier gene of *Drosophila*. *Development* **127**, 3971–3980 (2000).
 16. Kopp, A., Blackman, R. K. & Duncan, I. Wingless, decapentaplegic and EGF receptor signaling pathways interact to specify dorso-ventral pattern in the adult abdomen of *Drosophila*. *Development* **126**, 3495–3507 (1999).
 17. Abzhinov, A. & Kaufman, T. C. Crustacean (malacostracan) Hox genes and the evolution of the arthropod trunk. *Development* **127**, 2239–2249 (2000).
 18. Schutt, C. & Nothiger, R. Structure, function and evolution of sex-determining systems in dipteran insects. *Development* **127**, 667–677 (2000).
 19. Wulbeck, C. & Simpson, P. The expression of pannier and achaete-scute homologues in a mosquito suggests an ancient role of pannier as a selector gene in the regulation of the dorsal body pattern. *Development* **129**, 3861–3871 (2002).
 20. Majerus, M. E. N. *Ladybirds* (HarperCollins, London, 1994).
 21. Majerus, M. E. N. *Melanism: Evolution in Action* (Oxford Univ. Press, 1998).
 22. Madhavan, M. M. & Madhavan, K. Morphogenesis of the epidermis of adult abdomen of *Drosophila*. *J. Embryol. Exp. Morphol.* **60**, 1–31 (1980).
 23. Ashburner, M. *Drosophila. A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989).
 24. Kwiatkowski, J. & Ayala, F. J. Phylogeny of *Drosophila* and related genera: Conflict between molecular and anatomical analyses. *Mol. Phylogenet. Evol.* **13**, 319–328 (1999).
 25. Maddison, D. & Maddison, W. *MacClade 4: Analysis of Phylogeny and Character Evolution* (Sinauer, Sunderland, Massachusetts, 2000).

Acknowledgements We thank C. Nelson for help with fly collection; D. Lachaise and P. O'Grady for species identification; T. Markow, L. Andrew (Tucson Stock Center), J. Coyne and D. Lachaise for providing fly stocks; F. Laski and D. Godt for the Bab2 antibody; V. Kassner for invaluable technical assistance; A. Rokas for help with character reconstruction; and B. Williams, A. Kopp, A. Rokas and C. Nelson for discussions on the project. N.G. has been funded by the Howard Hughes Medical Institute and the Philippe foundation and is supported by an EMBO long-term post-doctoral fellowship. The project was supported by the Howard Hughes Medical Institute (S.B.C.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.B.C. (sbcarrol@facstaff.wisc.edu).

Regulatory evolution of *shavenbaby/ovo* underlies multiple cases of morphological parallelism

Elio Sucena^{1*}†, Isabelle Delon^{2*}, Isaac Jones³, François Payre² & David L. Stern¹

¹Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544, USA

²Centre de Biologie du Développement, Bâtiment 4R3, 118 Route de Narbonne, 31062 cedex4 Toulouse, France

³Vanguard High School, 317 East 67th Street, New York, New York 10021, USA

* These authors contributed equally to this work

† Present address: Instituto Gulbenkian de Ciência, Rua da quinta grande 6, 2780-156 Oeiras, Portugal

Cases of convergent evolution that involve changes in the same developmental pathway, called parallelism, provide evidence that a limited number of developmental changes are available to evolve a particular phenotype¹. To our knowledge, in no case are the genetic changes underlying morphological convergence understood. However, morphological convergence is not gener-

ally assumed to imply developmental parallelism². Here we investigate a case of convergence of larval morphology in insects and show that the loss of particular trichomes, observed in one species of the *Drosophila melanogaster* species group, has independently evolved multiple times in the distantly related *D. virilis* species group³. We present genetic and gene expression data showing that regulatory changes of the *shavenbaby/ovo* (*svb/ovo*) gene underlie all independent cases of this morphological convergence. Our results indicate that some developmental regulators might preferentially accumulate evolutionary changes and that morphological parallelism might therefore be more common than previously appreciated.

The pattern of microtrichiae (hereafter referred to as trichomes) on the ventral surface of first-instar larvae seems to be conserved across the genus *Drosophila*, whereas the dorsal and lateral surfaces have repeatedly evolved different patterns³. In most species, three kinds of trichome are produced in a specific pattern on the dorsal surface⁴. In a single species of the *D. melanogaster* subgroup, *D. sechellia*, many thin trichomes have been lost and the cells instead differentiate naked cuticle^{3,5}.

The patterning of dorsal trichomes is an interesting case of convergent morphological evolution, because four species from the *D. virilis* species group (*D. ezoana*, *D. borealis* eastern, *D. lacicola* and *D. montana*) also show evolutionary loss of the thin trichomes³ (Supplementary Fig. 1). All other species of the *D. virilis* species group that we have examined (*D. americana*, *D. borealis* western, *D. canadiana*, *D. flavomontana*, *D. kanekoi*, *D. littoralis*, *D. lummei*, *D. novamexicana* and *D. virilis*) and the more distantly related *D. arizonae* possess a lawn of trichomes similar to that observed in *D. melanogaster*. By mapping these phenotypes onto a recent molecular phylogeny of the *D. virilis* species group⁶, we can infer that at least three evolutionary transitions are required to explain the current distribution of trichome loss (Fig. 1). The convergence of trichome patterns in different fly lineages indicates that these changes might be driven by natural selection^{7,8}, although the selection pressure has not yet been identified. In addition, the evolutionary loss of trichomes in first-instar larvae mirrors an ontogenetic loss of the same trichomes in second-instar and third-instar larvae⁹ in all species we have examined, suggesting that these trichomes have a special function in first-instar larvae.

In theory, many genes might have evolved to alter the patterning of larval trichomes. For example, the *wingless* (*wg*) and *hedgehog* (*hh*) pathways and the *lines* gene are involved in patterning the trichomes on the dorsal epidermis^{4,10}. It might therefore be interesting to test whether evolution of patterning genes involved in segmentation can account for the evolution of trichome patterns. However, it has been shown³ that six genes involved in segmentation (*wg*, *gooseberry-distal*, *patched*, *engrailed*, *abdominal-A* and *hunchback*) are expressed identically in 12 species of the *D. virilis* species group, indicating that differences in trichome patterning might have evolved by changes in genes downstream of the segmentation pathway.

We have previously reported, after a full-genome genetic scan, that regulatory evolution at the *svb* gene accounts fully for the difference in trichome pattern between *D. sechellia* and other species of the *D. melanogaster* species group⁵. It has been shown¹¹ that the transcription factor Svb acts to switch cells between naked cuticle and the production of trichomes. In *D. melanogaster* embryos, *svb* is genetically required for trichome formation; when *svb* is expressed in a cell, that cell autonomously differentiates trichomes, whose morphology is determined by other patterning genes^{4,10}. Therefore *svb* integrates numerous sources of information (including the *wg*, *hh*, DER (for *Drosophila* epidermal growth factor receptor), homeotic and dorsal-ventral patterning systems) to specify the final pattern of trichomes.

To determine the genetic nature of the phenotypic differences observed in the *D. virilis* species group, we performed interspecific

genetic crosses. Although most *D. virilis* group species do not interbreed¹², we succeeded in obtaining interspecific hybrid larvae in three cases. First, we crossed *D. borealis* eastern females with *D. montana* males, both of which have a naked phenotype (Fig. 1 and 2a, b). Four larvae of unknown sex were recovered that all had the naked phenotype (Fig. 2c). This cross suggests that the naked phenotype is not caused by different autosomal recessive genes in each species. However, there are several possible genetic explanations for this result: the naked phenotype might be caused by a recessive X-linked gene (if all the larvae were male), by the same recessive X-linked gene in both species (if any of the larvae were female), by the same recessive autosomal gene in both species, or by a dominant allele in one or both of the species.

We succeeded in performing two further crosses that narrowed down these possibilities. We obtained larvae from a cross between *D. virilis* females (Fig. 2d) and *D. borealis* eastern males (Fig. 2a). The anterior halves of these larvae were mounted to analyse the trichome pattern, and DNA was prepared from the posterior half for a polymerase chain reaction assay. Between *D. virilis* and *D. borealis* eastern, we identified a restriction-site polymorphism in the *svb* gene, which is X-linked in *D. melanogaster* and *D. virilis*. Our DNA

analysis therefore allowed the determination of both larval sex and the origin of the *svb* allele(s). Two larvae carried only the *D. virilis* *svb* allele (males); two other larvae carried *svb* alleles from both species (females). All four larvae had a *D. virilis*-like trichome pattern (Fig. 2e), indicating that the *D. virilis* trichome pattern is dominant to the more naked *D. borealis* eastern pattern, just as it is in the *D. melanogaster* species group⁵. In a reciprocal cross between *D. borealis* eastern females and *D. virilis* males, two larvae were obtained. Both carried only the *D. borealis* *svb* allele (males) and had the *D. borealis* eastern pattern of trichomes (Fig. 2f), confirming the X-linkage of the factor(s) responsible for the difference in trichome pattern. These species carry multiple X-chromosome inversions¹³ that are likely to be responsible for the observed absence of recombination between heterospecific X chromosomes (A. Hoikala, personal communication), which prevented further genetic mapping of the factor(s) involved. Together, the interspecific genetic data show that the presence of trichomes is dominant to the absence of trichomes and that trichome pattern segregates with the X chromosome, which excludes about 80% of the genome but includes the *svb* gene.

Given the central role of *svb* in patterning trichomes in

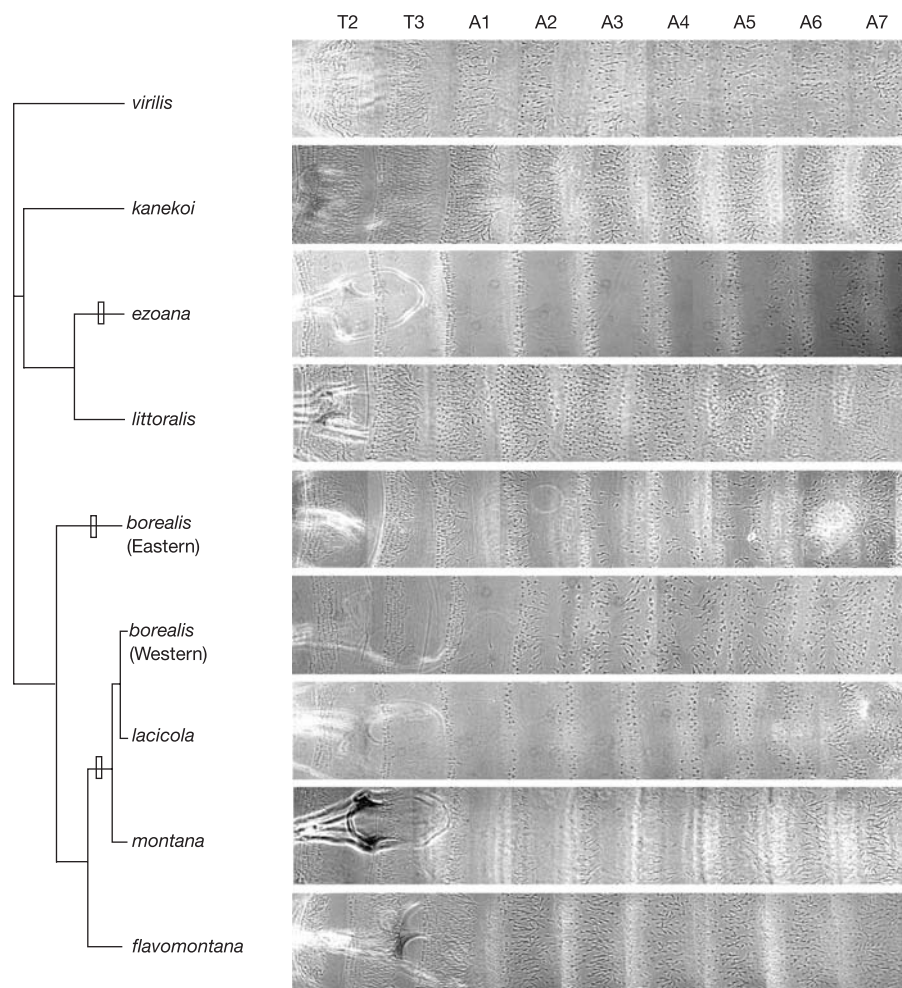


Figure 1 The pattern of dorsal trichomes has evolved repeatedly in the *D. virilis* species group. A phylogeny of the species (adapted from ref. 6) is shown to the left, with branch lengths proportional to molecular divergence. The oldest divergence event is estimated at about 11 Myr BP (ref. 6). The dorsal cuticle is shown for each species, with anterior to the left and the thoracic (T2 and T3) and abdominal (A1–A7) segments labelled at the top. Some species (e.g. *D. kanekoi*, *D. littoralis* and *D. flavomontana*) have an extensive lawn of fine trichomes in every segment, whereas others (*D. borealis* eastern and western,

D. lacicola and *D. montana*) have naked cuticle to different extents in more anterior segments. *D. ezoana* has the most extreme loss of fine trichomes in this species group. At least three evolutionary transitions are required to explain the distribution of naked and intermediate-naked phenotypes. One parsimonious mapping involves three evolutionary losses of trichomes, assuming that the presence of trichomes is ancestral, indicated by three open boxes on the branches of the phylogeny.

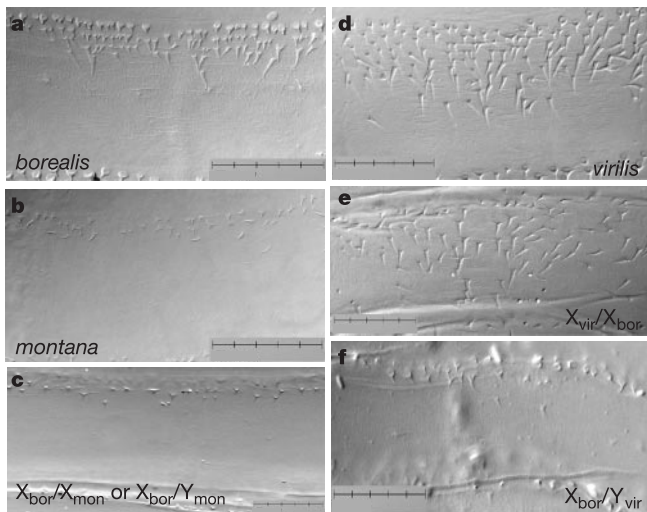


Figure 2 Dorsal first-abdominal segments of parental and species-hybrid larvae. **a, b**, *D. borealis* eastern (**a**) and *D. montana* (**b**) show a loss of the fine trichomes. **c**, A hybrid between a *D. borealis* eastern female and a *D. montana* male has the naked cuticle phenotype. The symbols indicate that these larvae carry either an X chromosome from each species (X_{bor}/X_{mon}) or an X chromosome from only *D. borealis* eastern (X_{bor}/Y_{mon}). **d**, *D. virilis* has fine trichomes over much of this segment. **e**, A hybrid between a female *D. virilis* and a male *D. borealis* eastern has the *D. virilis* pattern of trichomes. Because this larva carries an X chromosome from each species (X_{vir}/X_{bor}), this indicates that the *D. virilis* pattern is dominant to the *D. borealis* eastern pattern. **f**, A hybrid male between a female *D. borealis* eastern and a *D. virilis* male has the *D. borealis* eastern pattern of naked cuticle. This larva carries an X chromosome only from *D. borealis* eastern (X_{bor}/Y_{vir}), indicating the X-linked nature of the gene(s) causing the *D. borealis* eastern pattern. Scale bars, 10 μ m divisions.

D. melanogaster and its genetic co-segregation with trichome pattern in the *D. virilis* species group, we tested whether the evolution of divergent trichome patterns could be explained by the evolution of the *svb* expression pattern in embryos from nine

species from the *D. virilis* species group. We found that the embryonic pattern of *svb* expression closely matches the pattern of trichomes on the first-instar larval cuticle (Fig. 3). In species with an evolutionary loss of dorsal trichomes, *svb* transcription is absent from corresponding cells that differentiate naked cuticle. This suggests a simple model in which the transcriptional enhancer promoting *svb* expression in the domain that produces thin trichomes is turned off in some lineages.

When analysed in detail, however, the patterning of trichomes indicates a more complicated history of *svb* regulatory evolution. In *D. sechellia* and *D. ezoana*, essentially all thin trichomes are lost from all except the most posterior segments. In contrast, in the *D. montana* subphylad, species fall along a continuum from very hairy (*D. flavomontana*) to almost completely naked (*D. laticola*), with *D. borealis* eastern and *D. montana* having an intermediate naked phenotype (see Fig. 1). In the intermediate phenotypes, thin trichomes are absent to different degrees from the more anterior segments, whereas they are retained in the more posterior segments. These phenotypic differences might have been thought to imply that different genetic mechanisms were responsible for morphological evolution. However, we again observe a strict correlation between *svb* expression and the trichome patterns along the anterior–posterior axis (Fig. 4), indicating that regulatory changes in *svb* expression are involved in all of these evolutionary events. This might be due either to different changes in the *cis*-regulatory region of *svb* or potentially to different *trans*-acting changes in the different lineages. The latter possibility would imply the evolution of additional genes on the X chromosome controlling *svb* expression. This was genetically proved not to occur between *D. melanogaster* and *D. sechellia*⁵. However, rejecting this hypothesis for the *D. virilis* species group will require the identification and characterization of the appropriate regulatory regions of *svb* from these species.

We found that *svb* regulatory evolution underlies all observed cases of convergent evolution of the larval trichome pattern. In the epidermis, *svb* works like a morphogenetic switch¹¹, making it an effective point in the developmental cascade to generate alternative trichome patterns. Alterations of broader-acting patterning genes might cause pleiotropic consequences, whereas changes to indi-

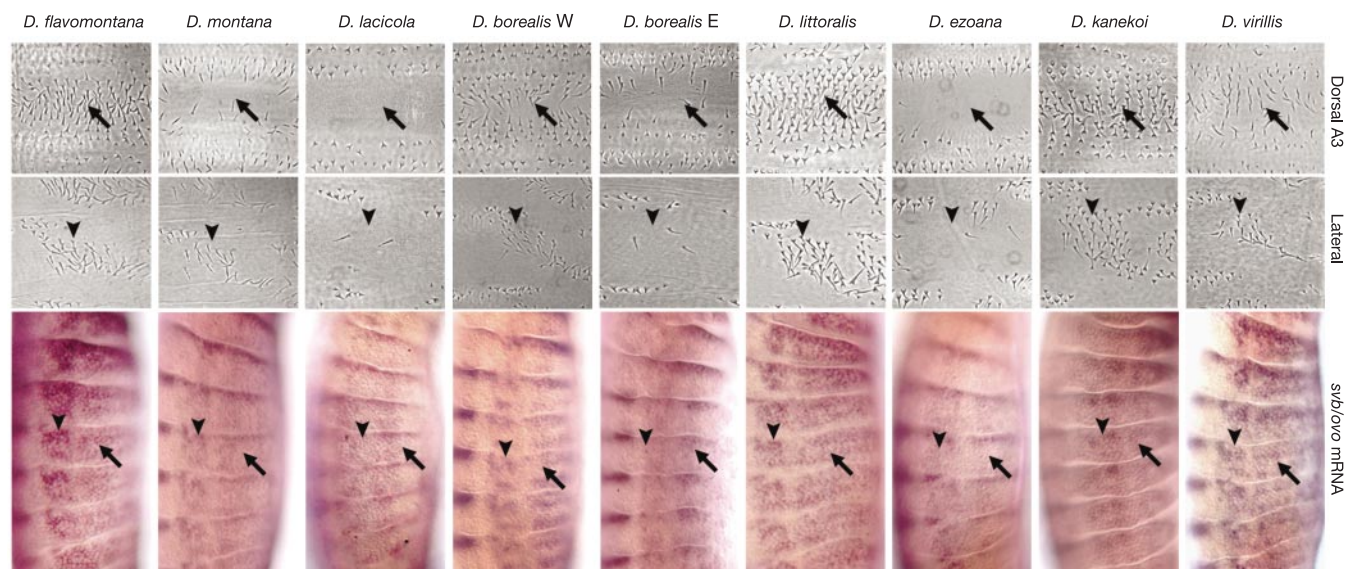


Figure 3 The expression pattern of *svb* is strictly correlated with the pattern of fine trichomes across the *D. virilis* group of species. The dorsal and lateral aspects of the third abdominal segment of first-instar larvae are shown in the first two columns. The embryonic pattern of *svb* expression, as revealed by *in situ* hybridization with a *D. virilis*

svb probe, is shown in the right-hand column. The regions of the dorsal and lateral fine hairs are indicated with arrows and arrowheads, respectively, in all panels. The hairs in the lateral panels for *D. borealis* eastern and *D. laticola* are innervated bristles whose production is not dependent on *svb* function.

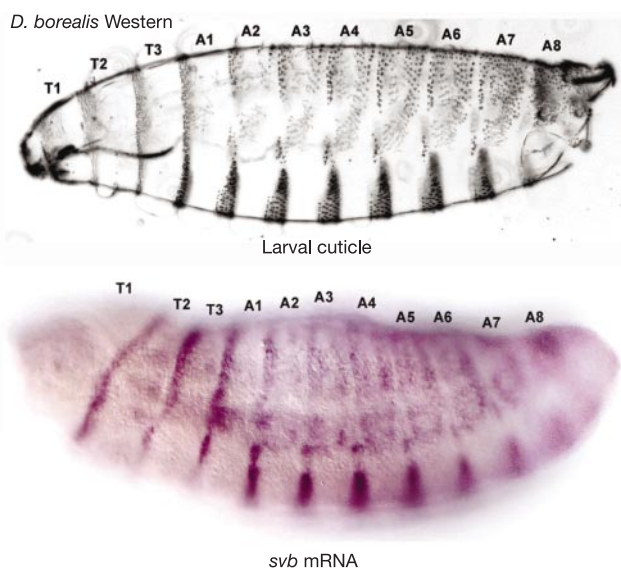


Figure 4 In *Drosophila borealis* western, which has an intermediate dorsal pattern of fine trichomes and naked cuticle, the expression of *svb* (lower panel) is strictly correlated with the pattern of trichomes (upper panel) along the anterior–posterior axis. The thoracic and abdominal segments are labelled.

vidual genes downstream of *svb*, presumably the cytoskeletal genes sculpting the apical extensions that form trichomes, might have no effect. In addition, the modular nature of *svb* regulatory regions permits alterations in part of the cuticular pattern without pleiotropic consequences⁵.

It has previously been recognized that organs that are identical by descent—homologous—need not be constructed by identical genetic mechanisms¹⁴. Our results indicate that changes in the same genetic mechanisms might result in morphological convergence^{15–17}. Therefore, observations of identical genetic mechanisms underlying similar morphologies do not necessarily imply homology: phylogenetic information must also be considered. □

Methods

Fly stocks

Flies were obtained from the Species Stock Center (stockcenter.ark.arizona.edu) and maintained on standard cornmeal agar medium: *D. americana* 15010-0951.1, *D. borealis* eastern 15010-0961.0, *D. borealis* western 15010-0961.3, *D. ezoana* 15010-0971.0, *D. flavomontana* 15010-0981.0, *D. kanekoi* 15010-1061.0, *D. laticola* 15010-0991.0, *D. littoralis* 15010-1001.0, *D. lummei* 15010-1011.1, *D. montana* 15010-1021.6, *D. novamexicana* 15010-1031.0 and *D. virilis* 15010-1051.83. Larval cuticles were prepared with standard protocols and photographed using phase-contrast and differential interference contrast microscopy on a Zeiss Axioplan.

Cloning of *svb/ovo* fragment

A fragment of the *svb/ovo* gene was cloned from *D. virilis* and *D. borealis* by polymerase chain reaction with primers for the conserved zinc-finger region: forward primer, 5'-CGGCCACGGCATCAARAAYCCNYT-3'; reverse primer, 5'-GACACTGTGCACCTTCTGGCAA-3'. We found an *Eco* RV restriction site difference between the *D. borealis* eastern and *D. virilis* alleles, which we used to identify the alleles present in hybrid larvae.

In situ hybridization

Plasmids containing a fragment of the *svb* gene from *D. virilis* were cut with appropriate restriction enzymes and used as templates for the preparation of digoxigenin-labelled sense and antisense RNA using Ambion kits. *In situ* hybridization was performed with standard procedures¹¹ on 0–24-h embryos collected at 22 °C and fixed with 37% formaldehyde.

Received 12 December 2002; accepted 22 April 2003; doi:10.1038/nature01768.

- Hodin, J. Plasticity and constraints in development and evolution. *Mol. Dev. Evol.* **288**, 1–20 (2000).
- Wittkopp, P. J., Williams, B. L., Selegue, J. E. & Carroll, S. B. *Drosophila* pigmentation evolution: Divergent genotypes underlying convergent phenotypes. *Proc. Natl Acad. Sci. USA* **100**, 1808–1813 (2003).

- Dickinson, W. J., Tang, Y., Schuske, K. & Akam, M. Conservation of molecular prepatterning during the evolution of cuticle morphology in *Drosophila* larvae. *Evolution* **47**, 1396–1406 (1993).
- Bokor, P. & DiNardo, S. The roles of *hedgehog*, *wingless* and *lines* in patterning the dorsal epidermis in *Drosophila*. *Development* **122**, 1083–1092 (1996).
- Sucena, E. & Stern, D. L. Divergence of larval morphology between *Drosophila sechellia* and its sibling species caused by cis-regulatory evolution of *ovo/shaven-baby*. *Proc. Natl Acad. Sci. USA* **97**, 4530–4534 (2000).
- Spicer, G. S. & Bell, C. D. Molecular phylogeny of the *Drosophila virilis* species group (Diptera: Drosophilidae) inferred from mitochondrial 12S and 16S ribosomal RNA genes. *Ann. Ent. Soc. Am.* **95**, 156–161 (2002).
- Darwin, C. *The Origin of Species by means of Natural Selection or The Preservation of Favored Races in the Struggle for Life* (Modern Library, New York, 1859).
- Endler, J. A. *Natural Selection in the Wild* (Monographs in Population Biology, ed. May, R. M.) (Princeton Univ. Press, 1986).
- Kuhn, D. T., Sawyer, M., Ventimiglia, J. & Sprey, T. E. Cuticle morphology changes with each larval molt in *D. melanogaster*. *Drosophila Inform. Serv.* **71**, 218–222 (1992).
- Hatini, V., Bokor, P., Goto-Mandeville, R. & DiNardo, S. Tissue- and stage-specific modulation of Wingless signaling by the segment polarity gene lines. *Genes Dev.* **14**, 1364–1376 (2000).
- Payre, F., Vincent, A. & Carreno, S. *ovo/ovb* integrates Wingless and DER pathways to control epidermis differentiation. *Nature* **400**, 271–275 (1999).
- Throckmorton, L. H. in *The Genetics and Biology of Drosophila* (eds Ashburner, M., Carson, H. L. & Thompson, J. N.) 227–296 (Academic, London, 1982).
- Hsu, T.-C. *Chromosomal Variation and Evolution in the virilis group of Drosophila*. Thesis, Univ. Texas (1951).
- Wray, G. A. & Abouheif, E. When is homology not homology? *Curr. Opin. Genet. Dev.* **8**, 675–680 (1998).
- Stewart, C.-B. & Wilson, A. C. Sequence convergence and functional adaptation of stomach lysozymes from foregut fermenters. *Cold Spring Harb. Symp. Quant. Biol.* **52**, 891–899 (1987).
- French-Constant, R. H. The molecular and population genetics of cyclodiene insecticide resistance. *Insect Biochem. Mol. Biol.* **24**, 335–345 (1994).
- Wittkopp, P. J., Vaccaro, K. & Carroll, S. B. Evolution of yellow gene regulation and pigmentation in *Drosophila*. *Curr. Biol.* **12**, 1547–1556 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank C. Braendle, M. Grbic, A. McGregor, H. Teotonio, L. Waltzer and members of the F.P. lab for comments on the manuscript. This work was supported by a grant from Association pour la Recherche sur le Cancer (ARC) to F.P., fellowships from Ministère de la Jeunesse, de l'Éducation Nationale et de la Recherche and ARC to I.D., and a BBSRC David Phillips Research Fellowship and NIH Grant to D.L.S.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.L.S. (dstern@princeton.edu). The NCBI accession numbers for *svb/ovo* fragments from *D. virilis* and *D. borealis* are AY184285 and AY184286, respectively.

Trans-synaptic shift in anion gradient in spinal lamina I neurons as a mechanism of neuropathic pain

Jeffrey A. M. Coull^{1,2}, Dominic Boudreau¹, Karine Bachand¹, Steven A. Prescott^{1,2}, Francine Nault^{1,2}, Attila Sik¹, Paul De Koninck¹ & Yves De Koninck^{1,2}

¹Neurobiologie Cellulaire, Centre de recherche Université Laval Robert-Giffard, Québec, Québec G1J 2G3, Canada

²Department of Pharmacology and Therapeutics, McGill University, Montréal, Québec H3G 1Y6, Canada

Modern pain-control theory¹ predicts that a loss of inhibition (disinhibition) in the dorsal horn of the spinal cord is a crucial substrate for chronic pain syndromes². However, the nature of the mechanisms that underlie such disinhibition has remained controversial^{3–6}. Here we present evidence for a novel mechanism of disinhibition following peripheral nerve injury. It involves a trans-synaptic reduction in the expression of the potassium–chloride exporter KCC2, and the consequent disruption of anion homeostasis in neurons of lamina I of the superficial dorsal horn, one of the main spinal nociceptive output pathways⁷. In our

experiments, the resulting shift in the transmembrane anion gradient caused normally inhibitory anionic synaptic currents to be excitatory, substantially driving up the net excitability of lamina I neurons. Local blockade or knock-down of the spinal KCC2 exporter in intact rats markedly reduced the nociceptive threshold, confirming that the reported disruption of anion homeostasis in lamina I neurons was sufficient to cause neuropathic pain.

Peripheral neuropathy was induced in rats by chronically constricting the sciatic nerve (Fig. 1a). Recent data indicate that injury to neurons can cause hyperexcitability (sensitization) through modification of the transmembrane anion gradient⁸, but it is not known whether this can occur trans-synaptically (downstream from the injured neuron). To test whether the hyperexcitability of superficial dorsal horn (SDH) neurons that follows peripheral nerve injury⁹ (PNI) is associated with a modification of the anion gradient (∇_{anion}), we measured the anion reversal potential (E_{anion}) by gramicidin-perforated patch-clamp recording. This technique avoids disrupting the intracellular anion concentration¹⁰. Responses to exogenous GABA (γ -aminobutyric acid) application showed that

E_{anion} of lamina I (LI) neurons taken from PNI rats was -49.0 ± 2.3 mV ($n = 9$), compared with -72.6 ± 3.5 mV ($n = 5$, $P < 0.005$) in LI neurons from naive rats (Fig. 1b–d). Resting membrane potential was not significantly different between PNI (-62 ± 4 mV, $n = 7$) and naive rat LI neurons (-61 ± 2 mV, $n = 16$, $P > 0.1$). Spontaneous and evoked postsynaptic currents (PSCs) recorded from PNI rat LI neurons in the presence of fast glutamate-receptor (GluR) blockers were also inward (depolarizing from rest), their mean reversal potential increasing by 16.1 mV relative to that in LI neurons from naive rats (PNI, $n = 6$; naive, $n = 4$).

We next tested whether other properties of synaptic transmission were altered in the SDH after PNI. Inhibitory miniature PSCs (mPSCs) in adult LI neurons from naive rats are mediated by glycine receptors (GlyRs) alone, despite GABA and glycine co-release from local inhibitory interneurons¹¹ (Fig. 2a). Although GluR-mediated mPSCs were unaffected by PNI (Fig. 2b), in all cells tested from PNI rats, a population of outward mPSCs at 0 mV persisted in the presence of the GlyR antagonist strychnine (up to $1 \mu\text{M}$; $n = 4$). These remaining mPSCs were mediated by GABA_A receptors, as they were blocked by bicuculline ($10 \mu\text{M}$) and had prolonged decay kinetics compared with the GlyR-mediated component ($\tau_{\text{D(GABA)}} = 34.0 \pm 2.9$ ms, $n = 5$; compared with $\tau_{\text{D(GlyR)}} = 11.3 \pm 1.3$ ms, $n = 6$, $P < 0.01$; Fig. 2c).

Kinetic analysis further revealed that the decay phase of $36.9 \pm 2.3\%$ of mPSCs followed a dual exponential function ($\tau_{\text{D1}} = 7.5 \pm 2.0$ ms and $\tau_{\text{D2}} = 51.3 \pm 7.9$ ms, $n = 6$; Fig. 2c). These events included both a GABA_A-receptor and a GlyR-mediated component, as either bicuculline or strychnine could lead to the abolition of their respective components ($n = 4$). Therefore, in parallel with the collapsed ∇_{anion} , PNI caused reorganization at LI synapses, thereby unmasking a GABA_A-receptor component. This synaptic organization is similar to that observed in immature LI–II neurons¹¹. The net effect of this synaptic switch is that it yields a population of quantal synaptic events with significantly longer decay kinetics.

To examine the function of the PNI-induced GABA_A-receptor-mediated contribution to mPSCs, we analysed both the peak conductance and the frequency of mPSCs. This was performed using CsCl-filled pipettes to clamp the E_{anion} at 0 mV in LI neurons taken from both PNI and naive rats, to prevent biased detection of mPSCs resulting from changes in driving force. Peak conductance of GlyR-only mPSCs recorded in LI neurons taken from PNI rats was significantly lower than (about half) that recorded from naive rat LI neurons (Fig. 2d). The addition of GABA_A-receptor-mediated events in the PNI condition, however, partially compensated for the decrease in GlyR-only conductance. The peak conductance of GluR-mediated quantal events did not differ significantly between LI neurons taken from naive rats and those from PNI rats (Fig. 2d).

Factoring together the changes in peak conductance, kinetics and driving force, the net charge carried by GlyR-only mPSCs at resting membrane potential in LI neurons from PNI rats was only just over one-third that in naive rats (Fig. 2e). With the contribution of GABA_A receptors, however, the net charge carried by mPSCs in PNI rats rose back to a level similar to that mediated by GlyRs in naive rats.

The frequency of GlyR-only mPSCs recorded in LI neurons from PNI rats was observed to be significantly less (0.13 ± 0.04 Hz, $n = 5$) than that of GlyR-only mPSCs in naive rat LI neurons (0.18 ± 0.04 Hz, $n = 6$, $P < 0.05$; Fig. 2f). As with peak conductance, however, the addition of the GABA_A-receptor-mediated mPSCs compensated for the PNI-induced decrease in frequency (0.22 ± 0.10 Hz, $n = 4$, $P > 0.5$ for all GABA_A-receptor and/or GlyR-mediated events combined). By contrast, there was no significant change in the frequency of GluR-mediated events in LI neurons isolated from PNI rats (1.51 ± 0.90 Hz, $n = 9$) compared with naive LI neurons (0.82 ± 0.40 Hz, $n = 5$, $P > 0.3$; Fig. 2f).

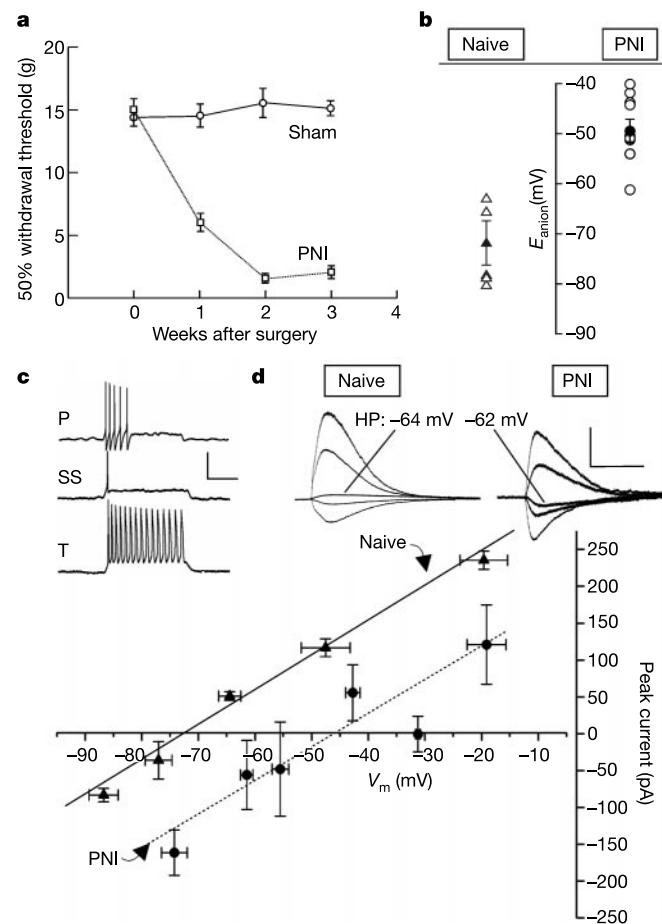


Figure 1 PNI induced a collapse of ∇_{anion} in LI neurons in the ipsilateral SDH. **a**, PNI ($n = 23$), but not sham surgery ($n = 11$), caused a significant reduction in the 50% nociceptive withdrawal threshold to mechanical stimulation of the hind paw in rats ($P < 0.01$). **b**, E_{anion} recorded from LI neurons of naive (open triangles) and PNI (open circles) rats. Solid symbol, mean $E_{\text{anion}} \pm \text{s.e.m.}$ **c**, All classes of LI neuron (that is, with phasic (P), single-spike (SS) and tonic (T) firing properties³⁰) showed a shift in E_{anion} in response to PNI. Scale (y , x), 50 mV, 150 ms. **d**, Mean peak current measured in LI neurons from naive (filled triangles) and PNI (filled circles) rats in response to applied GABA at various values of V_m . Horizontal standard-error bars represent differences between neurons in recording-pipette offset. Inset, representative raw traces. Scale (y , x), 0.6 nA, 1.0 s.

Thus, the result of these synaptic analyses revealed a shift from GlyR towards GABA_A-receptor-dominated synaptic transmission after PNI.

If depolarizing GABA_A-receptor/GlyR-mediated postsynaptic currents exert a net excitatory influence on PNI LI neurons, they

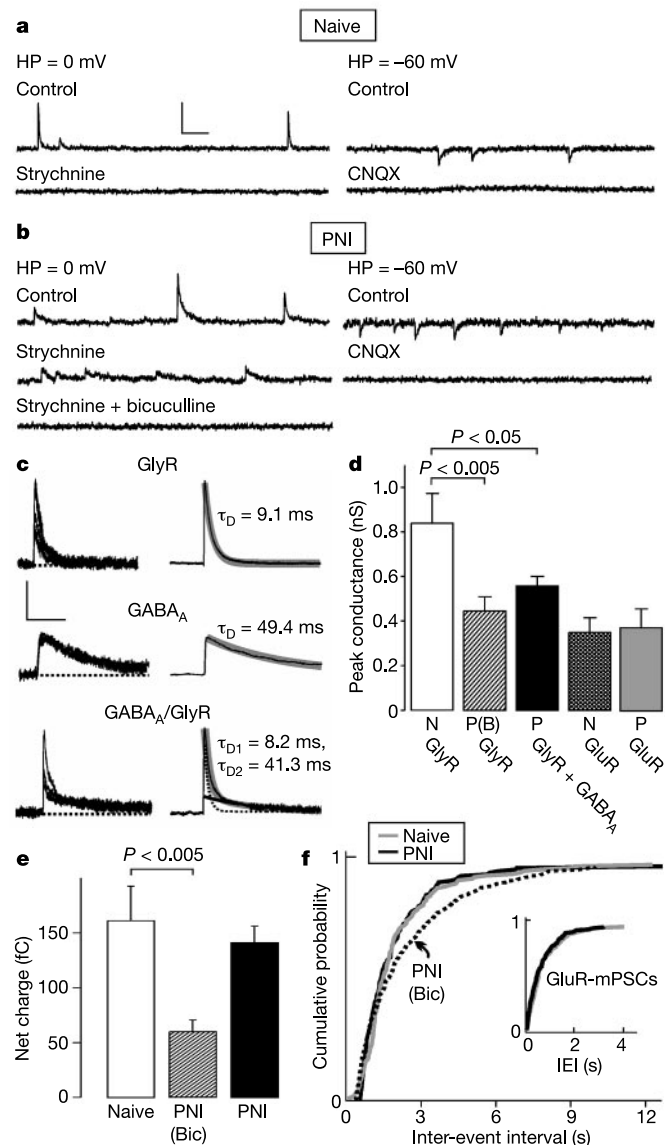


Figure 2 Switch from GlyR-only to mixed GABA_A-receptor and GlyR-mediated mPSCs following PNI in LI neurons. **a, b**, Pharmacological identification of outward (left) and inward (right) miniature synaptic events recorded at 0 mV and -60 mV, respectively, from naive (**a**) and PNI (**b**) rat LI neurons. Note the appearance of bicuculline-sensitive mPSCs in PNI rats. HP, holding potential. Scale (y, x), 40 pA, 300 ms. **c**, Left, superimposed individual mPSCs recorded from PNI rat LI neurons. GlyR-only, GABA_A-receptor-only and mixed GABA_A-receptor/GlyR-mediated PSCs were clearly identifiable by their sensitivity to strychnine and/or bicuculline, as previously described¹¹ (see also main text). Right, averages of more than 100 GlyR and GABA_A-receptor-mediated mPSCs recorded from a PNI rat LI neuron. Scale (y, x), 15 pA, 20 ms. **d**, Mean peak conductance of mPSCs recorded from naive (N; $n = 10$ for GlyR, $n = 5$ for GluR) and PNI (P; $n = 9$ for GlyR, $n = 8$ for GluR) LI neurons. P(B) indicates GlyR-mediated mPSCs recorded in PNI rat LI neurons ($n = 12$) at 0 mV in the presence of bicuculline. **e**, Net charge carried by GlyR-mediated mPSCs in naive rats ($n = 6$), by bicuculline-isolated GlyR-mediated mPSCs in PNI rats (PNI(Bic); $n = 4$), and by mixed GABA_A-receptor/GlyR-mediated mPSCs in PNI rat LI neurons (PNI; $n = 12$). **f**, Cumulative probability plot of isolated GlyR-only mPSC inter-event interval (IEI) in naive LI neurons compared with PNI rat LI neurons (PNI(Bic)). Inset, no effect of PNI on the frequency of GluR-mediated mPSCs.

should directly evoke action potentials and consequently lead to Ca^{2+} influx¹². To test this hypothesis, we first employed Ca^{2+} imaging of Fura-2-AM-loaded LI neurons in slices to obtain a large data set. Administration of exogenous GABA to neuronal somata caused a significant increase in the concentration of intracellular Ca^{2+} ($[\text{Ca}^{2+}]_i$) in 19% of LI neurons ($n = 53$; Fig. 3a, c) lying ipsilateral to the site of PNI. This represents a sevenfold increase compared with LI neurons found in naive and/or contralateral dorsal horn (where only 1 of 37 neurons showed an increase in $[\text{Ca}^{2+}]_i$; Fig. 3b, c). These responses were blocked by bicuculline ($10 \mu\text{M}$; $n = 5$) and by the voltage-sensitive Na^+ -channel blocker

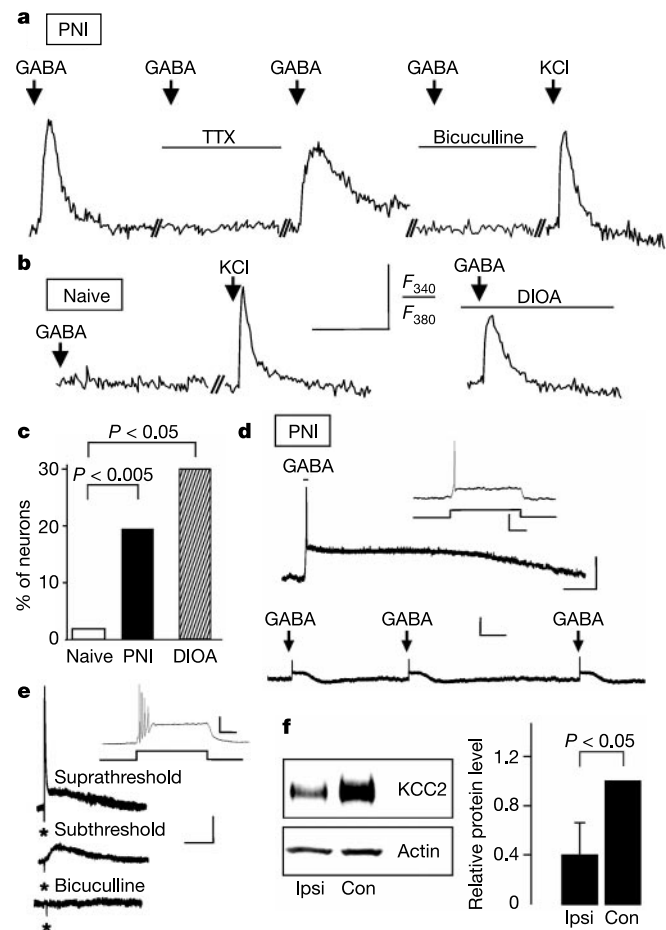


Figure 3 PNI-induced downregulation of KCC2 in SDH LI neurons ipsilateral to PNI led to GlyR/GABA_A-receptor-mediated excitation. **a**, Brief GABA application caused a TTX- and bicuculline-sensitive rise in $[\text{Ca}^{2+}]_i$ in a LI neuron from a Fura-2-AM-loaded slice from a PNI rat. **b**, KCl but not GABA application caused an increase in $[\text{Ca}^{2+}]_i$ in a naive rat LI neuron. In the presence of the KCC2-specific antagonist DIOA, however, GABA application did elicit a rise in $[\text{Ca}^{2+}]_i$ in a naive rat LI neuron. Scale (y, x), 0.02, 10 s. **c**, Compared with control naive rats, the proportion of LI neurons displaying a GABA-evoked increase in $[\text{Ca}^{2+}]_i$ was significantly increased after PNI ($\chi^2_{\text{corrected}} = 3.91$) or in the presence of DIOA ($\chi^2_{\text{corrected}} = 4.43$). **d**, Brief GABA application could repeatedly elicit action potentials in a LI neuron from a PNI rat. Upper scale (y, x), 5 mV, 200 ms; lower scale (y, x), 30 mV, 4 s. Inset, response to a depolarizing pulse illustrating that this was a single-spike neuron³⁰. Scale (y, x), 20 mV, 300 ms. **e**, Focal electrical stimuli (in the presence of CNQX/AP5) could elicit bicuculline-sensitive monosynaptic excitatory postsynaptic potentials in a PNI rat LI neuron. Scale (y, x), 5 mV, 250 ms. Inset, response to a depolarizing pulse showing that this was a phasic neuron³⁰. Scale (y, x), 20 mV, 300 ms. **f**, Left, immunoblotting revealed that KCC2 levels were decreased in the lumbar SDH lying ipsilateral (Ipsi), but not contralateral (Con), to the site of the PNI. Right, quantification of KCC2 protein levels (normalized to actin) measured from immunoblots ($n = 4$) as on the left (Ipsi normalized to Con).

tetrodotoxin (TTX; 1 μ M; $n = 31$). We then further confirmed electrophysiologically that applied GABA and synaptically elicited anionic postsynaptic potentials could directly evoke action potentials (Fig. 3d, e). These results indicate that postsynaptic anion fluxes can cause net excitation in LI neurons in PNI rats.

What cellular mechanism underlies this shift in V_{anion} ? Recent studies at early developmental stages in the brain have shown a link between the expression of the potassium–chloride exporter KCC2 and neuronal $[\text{Cl}^-]_i$ ^{13,14}. We therefore compared KCC2 protein levels by immunoblotting on horizontal slices of SDH. The KCC2 expression level in the lumbar SDH ipsilateral to the PNI was significantly reduced, to about half that on the side contralateral to the injury (Fig. 3f). In naive rats, there was no significant difference between the two sides ($n = 3$; not shown).

If a decrease in the expression of the KCC2 exporter leads to an increase in neuronal $[\text{Cl}^-]_i$ and, in turn, GABA_A-receptor/GlyR-mediated depolarization, a pharmacological blockade of the KCC2 exporter in LI neurons from naive rats should have the same effect. To test for this possibility, we bath-applied the selective KCC2 blocker [(dihydroindenyl)oxy]alkanoic acid (DIOA; 30 μ M) to naive spinal slices. As in the PNI condition, GABA application in the presence of DIOA caused an increase in $[\text{Ca}^{2+}]_i$ in 30% of naive LI neurons tested (Fig. 3b, c).

Although we have shown that, after PNI, GABA_A receptors and GlyRs mediated postsynaptic depolarization, the predictive value of miniature and evoked PSCs for overall neuronal excitability is

limited. This stems from the distinct input–output properties of neurons in intact animals, resulting from the many-fold greater frequency and magnitude of spontaneous PSCs bombarding cells *in vivo*¹⁵. Therefore, to assess whether the empirically determined changes in GABA_A-receptor/GlyR-mediated postsynaptic control were sufficient to account for the central sensitization that is known to follow PNI⁹, we simulated *in vivo* conditions using a biophysically realistic neuron model (see Supplementary Fig. 1). The simulation confirmed that, after PNI, the extent of LI neuronal sensitization varied as a function of E_{anion} , ranging from slight disinhibition to a net hyperexcitation.

To test whether this predicted hyperexcitability (sensitization) would result in a decrease in the stimulus threshold to evoke a nociceptive withdrawal reflex, we administered DIOA (15–30 μ g) directly to the lumbar enlargement of the spinal cord in intact rats through an intrathecal catheter. DIOA caused a rapid and reversible decrease in nociceptive threshold to both mechanical and thermal stimuli (Fig. 4a, b). This effect of DIOA did not occur through an alteration of anion conductance (see Supplementary Fig. 2). A similar decrease in nociceptive threshold was elicited through selective knock-down of the exporter using spinal administration of an antisense oligodeoxynucleotide against KCC2 messenger RNA (Fig. 4c), further confirming the functional impact of KCC2 downregulation.

The results show that the neuropathic pain that follows PNI can be explained by a downregulation of the KCC2 exporter and a resultant shift in the V_{anion} in spinal LI neurons. They also demonstrate that such a modification of V_{anion} in adult animals can occur in a neuron trans-synaptic to an injury site. Our findings thus provide a new avenue to understand disinhibition in nociceptive pathways. The inversion in polarity of the GABA_A-receptor/GlyR-mediated postsynaptic action forces a reinterpretation of data related to changes in the efficacy of GABAergic transmission, and provides a potential mechanistic basis for many of the reported consequences of central sensitization¹⁶. It is interesting to note, in this context, that LI nociceptive-specific neurons have been shown to respond to innocuous input when sensitized¹⁷.

A crucial feature of the spinal cord is that it uses two very distinct GABAergic inhibitory mechanisms: GABAergic control of the central terminals of sensory fibres already involves a depolarizing mechanism¹⁸, in contrast to dorsal horn cells, in which GABAergic inhibition involves hyperpolarization. Thus, the change in KCC2 expression reported here affects the polarity of GABA action in only one of the two inhibitory mechanisms controlling sensory input. This is confirmed by the fact that primary afferents lack expression of KCC2 (Fig. 4d, e; see Supplementary Fig. 3). This might explain the seemingly contradictory anti-allodynic effect of intrathecally applied GABA_A-receptor agonists^{19,20}.

GABA/glycine-mediated depolarization might also serve as a gating mechanism to enable excitation through voltage-sensitive Ca^{2+} channels and NMDA (*N*-methyl-D-aspartate) receptors²¹. Ca^{2+} influx through these channels is thought to be crucial for the sensitization of spinal neurons²². Indeed, blocking these Ca^{2+} channels in humans by drugs such as gabapentin and ketamine has proved to be highly efficacious in the treatment of neuropathic pain^{23–25}. However, the use of Ca^{2+} -channel blockers, particularly ketamine and other NMDA antagonists, is associated with many undesirable side effects^{25,26}. Our identification of the pivotal role of appropriate anion homeostasis in SDH neurons in maintaining adequate nociceptive gate control opens new avenues for the development of more selective therapeutic strategies for the management of chronic neuropathic-pain syndromes. □

Methods

Nerve injury

Briefly, PNI was induced by surgically implanting a polyethylene cuff (~2 mm in length, inner diameter 0.7 mm) around the sciatic nerve of adult male Sprague–Dawley rats as

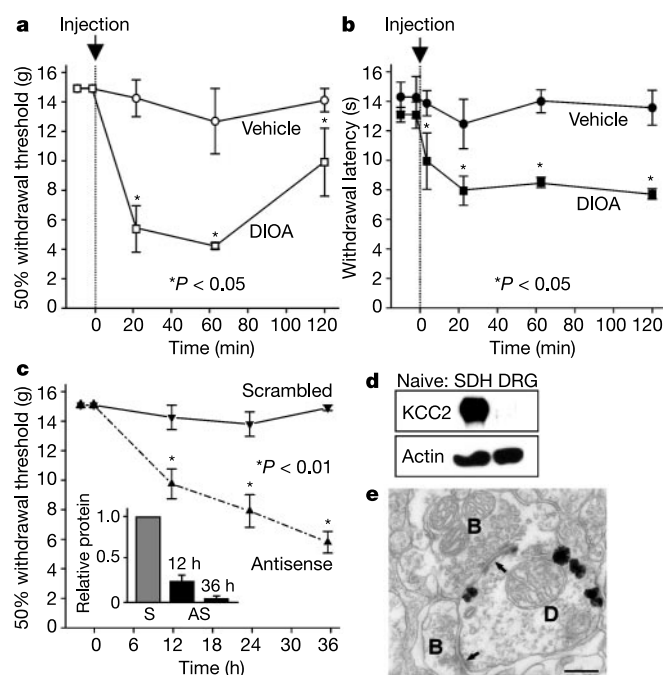


Figure 4 Selective blockade or knock-down of the postsynaptic KCC2 exporter in the SDH significantly reduced nociceptive threshold. **a**, Tactile nociceptive withdrawal threshold as a function of time after intrathecal injections of DIOA ($n = 5$) or vehicle ($n = 3$). **b**, Thermal nociceptive withdrawal latency as a function of time after intrathecal injections of DIOA ($n = 3$) or vehicle ($n = 3$). On withdrawal, the rats often licked their paw, indicating nociception. **c**, A significant decrease in the tactile nociceptive withdrawal threshold in naive rats was observed after local lumbar spinal (intrathecal) administration of a KCC2 antisense ($n = 8$), but not a scrambled ($n = 7$), oligodeoxynucleotide (0 h, 12 h and 24 h). Inset, decrease in spinal KCC2 protein levels (measured by immunoblots) after antisense (AS; 12 h or 36 h) or scrambled (S; 36 h) oligodeoxynucleotide treatment. **d**, Lack of KCC2 immunoreactivity in dorsal root ganglia (DRG) in a naive rat, compared with SDH. **e**, Electron micrograph illustrating the selective expression of KCC2 in SDH dendrites (D), but not synaptic boutons (B) (for quantitative details, see Supplementary Fig. 3). Arrows point to synapses. Scale, 0.2 μ m.

previously described²⁷. A group of rats also received sham surgery. Only animals that showed a gradual decrease in mechanical threshold (over 14–17 days) down to 2.0 g or less were used for further experiments.

Behavioural testing

Thermal and mechanical threshold for nociceptive withdrawal reflexes were tested as previously described²⁸.

Slice preparation

Parasagittal slices (300–350 μm) of spinal cord were prepared from adult (>50 days old) male rats as previously described¹¹. Slices were continually superfused (2–3 ml min^{-1}) with artificial cerebrospinal fluid (ACSF) containing (in mM): 126 NaCl, 26 NaHCO_3 , 10 glucose, 2.5 KCl, 2 CaCl_2 , 2 MgCl_2 , 1.25 NaH_2PO_4 , 0.001 TTX (bubbled with 95% $\text{O}_2/5\%$ CO_2 , $\sim\text{pH}$ 7.4); when measuring GABA_A-receptor/GlyR-mediated currents, 10 μM 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX) and 40 μM D-2-amino-5-phosphonopentanoate (AP5) were added to block fast glutamatergic neurotransmission.

Recordings

For perforated-patch recordings, the pipette tip was filled with a solution containing (in mM): 130 caesium gluconate, 5 CsCl, 2 MgCl_2 , 11 BAPTA, 1 CaCl_2 , 4 ATP, 0.4 GTP, 10 HEPES ($\sim\text{pH}$ 7.4). The pipette was back-filled with this same solution supplemented with 25 $\mu\text{g ml}^{-1}$ gramicidin D (gramicidin stock was at 10 mg ml^{-1} in dimethylsulphoxide (DMSO)). Recordings in this mode were selected when access resistance was stable between 25–45 M Ω . For whole-cell voltage-clamp recordings, pipettes were filled with the above solution without gramicidin D. For whole-cell current-clamp recordings, pipettes were filled with the same intracellular solution as for voltage clamp, except that potassium methyl sulphate (KMeSO_4) was used instead of caesium gluconate. To clamp E_{anion} at 0 mV, caesium gluconate was replaced with 110 mM CsCl in the intracellular solution. All whole-cell recordings at $E_{\text{anion}} = 0$ mV were made at a membrane potential, V_m , of -60 mV in the presence of GluR blockers. GABA was applied locally for 30–250 ms by pressure ejection through a patch micropipette. Data acquisition and analysis of PSCs were performed as previously described¹¹. All measurements are given as means $\pm$ s.e.m., except where indicated. Statistical significance was tested using Student's *t*-tests for comparison of mean values, chi-squared tests for contingency tables, and mixed-design analyses of variance (post-hoc Tukey's HSD test) for repeated measures.

Calcium imaging

Slices were prepared as detailed above. After a 15-min incubation in ACSF, slices were loaded with 10 μM Fura-2-AM in HEPES-buffered saline (+10% DMSO) for 1 h. Slices were washed for ~ 15 min with ACSF before being mounted in the recording chamber, where they were superfused with ACSF (2–3 ml min^{-1}). $[\text{Ca}^{2+}]_i$ was fluorometrically measured using a $\times 40$ water-immersion objective on a Zeiss Axioscope equipped with epifluorescence optics. Images were acquired using a TILL Photonics monochromator coupled to a CCD camera, and regions of interest (for ratioing) were drawn on clearly distinct neuronal cell bodies.

Immunoblotting

Horizontal slices (150 μm) of the SDH were made from the lumbar enlargement of both PNI and naive adult rats. Tissue extracts were prepared by homogenizing the slices with a Teflon pestle in a buffer containing 0.32 M sucrose, 0.5 mM Tris-HCl, pH 7.5, 2 mM EDTA, 2.5 mM β -mercaptoethanol, and a cocktail of protease inhibitors (Complete, Roche Diagnostics). Supernatants from 3,000g (20 min) and 10,000g (30 min) centrifugations were collected. Equal amounts of protein (20 μg per lane) diluted in sample buffer were preheated at 37 $^\circ\text{C}$ for 30 min, resolved by SDS-PAGE, and electroblotted onto nitrocellulose membranes. Membranes were blocked for 30 min in 5% non-fat dry milk in TBST buffer (150 mM NaCl, 10 mM Tris-HCl, pH 7.4, 0.05% Tween-20) and incubated overnight at 4 $^\circ\text{C}$ with a rabbit anti-KCC2 antibody (1:1,000, Upstate Biotechnology). After several washes in TBST, membranes were incubated for 30 min at room temperature with peroxidase-labelled goat anti-rabbit antibody (1:2,000). Chemiluminescent bands were detected using Super Signal Femto (Pierce Biotechnology). Digital images were captured with the VersaDoc imaging system and analysed with Quantity One software (BioRad).

Oligodeoxynucleotides

KCC2 antisense and scrambled oligodeoxynucleotides, phosphorothioated at all positions, were designed as previously described²⁹: antisense, 5'-TCTCCTGGGATTG CCGTCA-3' (+59 relative to the initial ATG); scrambled, 5'-TCTTCTTGAGACTGCAG TCA-3'.

Intrathecal injections

At least three days before drug administration, rats were anaesthetized with sodium pentobarbital (65 mg kg^{-1}), and a lumbar spinal catheter was inserted into the intrathecal space, as previously described²². On recovery from surgery, lower-body paralysis was induced through intrathecal lidocaine (2%, 30 μl) injection to confirm proper catheter localization. Only animals exhibiting appropriate, transient paralysis to lidocaine, as well as a lack of motor deficits, were used for behavioural testing. Following drug/vehicle administration, animals were killed and their vertebral column dissected to visually confirm correct placement of the catheter. Drugs included DIOA (10–30 μg , in 0.9% NaCl, 10% DMSO) and oligodeoxynucleotides (single doses of 2 μg at 0 h, 12 h and 24 h; 0.9% NaCl). Behavioural testing was performed as above (under blind conditions); normal (~ 15 g) mechanical threshold for withdrawal responses was confirmed in naive rats before receiving drug or vehicle. At the doses used, none of the compounds produced motor

disturbances or sedation, as assessed by grasping, righting and placing reflexes and behavioural observations²⁸.

Received 23 December 2002; accepted 18 June 2003; doi:10.1038/nature01868.

- Melzack, R. & Wall, P. D. Pain mechanisms: a new theory. *Science* **150**, 971–979 (1965).
- Woolf, C. J. & Salter, M. W. Neuronal plasticity: increasing the gain in pain. *Science* **288**, 1765–1769 (2000).
- Kontinen, V. K., Stanfa, L. C., Basu, A. & Dickenson, A. H. Electrophysiologic evidence for increased endogenous GABAergic but not glycinergic inhibitory tone in the rat spinal nerve ligation model of neuropathy. *Anesthesiology* **94**, 333–339 (2001).
- Moore, K. A. *et al.* Partial peripheral nerve injury promotes a selective loss of GABAergic inhibition in the superficial dorsal horn of the spinal cord. *J. Neurosci.* **22**, 6724–6731 (2002).
- Somers, D. L. & Clemente, F. R. Dorsal horn synaptosomal content of aspartate, glutamate, glycine and GABA are differentially altered following chronic constriction injury to the rat sciatic nerve. *Neurosci. Lett.* **323**, 171–174 (2002).
- Polgar, E. *et al.* Selective loss of spinal GABAergic or glycinergic neurons is not necessary for development of thermal hyperalgesia in the chronic constriction injury model of neuropathic pain. *Pain* **104**, 229–239 (2003).
- Light, A. R. *The Initial Processing of Pain* (Karger, Basel, 1992).
- van den Pol, A. N., Obrietan, K. & Chen, G. Excitatory actions of GABA after neuronal trauma. *J. Neurosci.* **16**, 4283–4292 (1996).
- Pitcher, G. M. & Henry, J. L. Cellular mechanisms of hyperalgesia and spontaneous pain in a spinalized rat model of peripheral neuropathy: changes in myelinated afferent inputs implicated. *Eur. J. Neurosci.* **12**, 2006–2020 (2000).
- Ebihara, S., Shirato, K., Harata, N. & Akaïke, N. Gramicidin-perforated patch recording: GABA response in mammalian neurones with intact intracellular chloride. *J. Physiol. (Lond.)* **484**, 77–86 (1995).
- Keller, A. F., Coull, J. A., Chery, N., Poisbeau, P. & de Koninck, Y. Region-specific developmental specialization of GABA-glycine cosynapses in laminae I–II of the rat spinal dorsal horn. *J. Neurosci.* **21**, 7871–7880 (2001).
- Nabekura, J. *et al.* Reduction of KCC2 expression and GABA_A receptor-mediated excitation after *in vivo* axonal injury. *J. Neurosci.* **22**, 4412–4417 (2002).
- Hubner, C. A. *et al.* Disruption of KCC2 reveals an essential role of K–Cl cotransport already in early synaptic inhibition. *Neuron* **30**, 515–524 (2001).
- Ganguly, K., Schinder, A. F., Wong, S. T. & Poo, M. GABA itself promotes the developmental switch of neuronal GABAergic responses from excitation to inhibition. *Cell* **105**, 521–532 (2001).
- Bernander, O., Douglas, R. J., Martin, K. A. & Koch, C. Synaptic background activity influences spatiotemporal integration in single pyramidal cells. *Proc. Natl Acad. Sci. USA* **88**, 11569–11573 (1991).
- Cook, A. J., Woolf, C. J., Wall, P. D. & McMahon, S. B. Dynamic receptive field plasticity in rat spinal cord dorsal horn following C-primary afferent input. *Nature* **325**, 151–153 (1987).
- Blomqvist, A. & Craig, A. D. Is neuropathic pain caused by the activation of nociceptive-specific neurons due to axonomic sprouting in the dorsal horn? *J. Comp. Neurol.* **428**, 1–4 (2000).
- Rudomin, P. & Schmidt, R. F. Presynaptic inhibition in the vertebrate spinal cord revisited. *Exp. Brain Res.* **129**, 1–37 (1999).
- Hwang, J. H. & Yaksh, T. L. The effect of spinal GABA receptor agonists on tactile allodynia in a surgically-induced neuropathic pain model in the rat. *Pain* **70**, 15–22 (1997).
- Malan, T. P., Mata, H. P. & Porreca, F. Spinal GABA_A and GABA_B receptor pharmacology in a rat model of neuropathic pain. *Anesthesiology* **96**, 1161–1167 (2002).
- Staley, K. J., Soldo, B. L. & Proctor, W. R. Ionic mechanisms of neuronal excitation by inhibitory GABA_A receptors. *Science* **269**, 977–981 (1995).
- Coderre, T. J. & Melzack, R. The contribution of excitatory amino acids to central sensitization and persistent nociception after formalin-induced tissue injury. *J. Neurosci.* **12**, 3665–3670 (1992).
- Serpell, M. G. Gabapentin in neuropathic pain syndromes: a randomised, double-blind, placebo-controlled trial. *Pain* **99**, 557–566 (2002).
- Rabben, T., Skjelbred, P. & Oye, I. Prolonged analgesic effect of ketamine, an *N*-methyl-D-aspartate receptor inhibitor, in patients with chronic pain. *J. Pharmacol. Exp. Ther.* **289**, 1060–1066 (1999).
- Martin, T. J. & Eisenach, J. C. Pharmacology of opioid and nonopioid analgesics in chronic pain states. *J. Pharmacol. Exp. Ther.* **299**, 811–817 (2001).
- Farber, N. B., Jiang, X. P., Heinkel, C. & Nemmers, B. Antiepileptic drugs and agents that inhibit voltage-gated sodium channels prevent NMDA antagonist neurotoxicity. *Mol. Psychiatry* **7**, 726–733 (2002).
- Mosconi, T. & Kruger, L. Fixed-diameter polyethylene cuffs applied to the rat sciatic nerve induce a painful neuropathy: ultrastructural morphometric analysis of axonal alterations. *Pain* **64**, 37–57 (1996).
- Coderre, T. J. & Van, E. I. The utility of excitatory amino acid (EAA) antagonists as analgesic agents. I. Comparison of the antinociceptive activity of various classes of EAA antagonists in mechanical, thermal and chemical nociceptive tests. *Pain* **59**, 345–352 (1994).
- Rivera, C. *et al.* The K⁺/Cl[−] co-transporter KCC2 renders GABA hyperpolarizing during neuronal maturation. *Nature* **397**, 251–255 (1999).
- Prescott, S. A. & de Koninck, Y. Four cell types with distinctive membrane properties and morphologies in lamina I of the spinal dorsal horn of the adult rat. *J. Physiol. (Lond.)* **539**, 817–836 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements Supported by Canadian Institutes of Health Research (CIHR) grants to Y.D.K. and A.S., and a grant from the Burroughs Wellcome Fund (BWF) to P.D.K. J.A.M.C. is the recipient of a CIHR doctoral award. P.D.K. is the recipient of a career award from the BWF. Y.D.K. and A.S. are scholars of the Fonds de la recherche en santé du Québec (FRSQ). We thank C. M. Cahill for help with some of the Ca^{2+} measurements, L. Pellegrini for advice on the quantification of immunoblotting, and O. Szalay for technical assistance. We also thank F. J. Alvarez, M. Deschênes, P. Drapeau and M. W. Salter for critical comments on the manuscript.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Y.D.K. (Yves.DeKoninck@crulrg.ulaval.ca).

Different domains of synaptotagmin control the choice between kiss-and-run and full fusion

Chih-Tien Wang¹, Juu-Chin Lu², Jihong Bai³, Payne Y. Chang³, Thomas F. J. Martin⁴, Edwin R. Chapman¹ & Meyer B. Jackson¹

¹Department of Physiology, ²Department of Comparative Biosciences, Endocrinology-Reproductive Physiology Program, ³Biophysics Program, and ⁴Department of Biochemistry, University of Wisconsin-Madison, 1300 University Avenue, Madison, Wisconsin 53706, USA

Exocytosis—the release of the contents of a vesicle—proceeds by two mechanisms^{1–6}. Full fusion occurs when the vesicle and plasma membranes merge. Alternatively, in what is termed kiss-and-run, vesicles can release transmitter during transient contacts with the plasma membrane. Little is known at the molecular level about how the choice between these two pathways is regulated. Here we report amperometric recordings of catecholamine efflux through individual fusion pores. Transfection with synaptotagmin (Syt) IV increased the frequency and duration of kiss-and-run events, but left their amplitude unchanged. Endogenous Syt IV, induced by forskolin treatment, had a similar effect. Full fusion was inhibited by mutation of a Ca^{2+} ligand in the C₂A domain of Syt I; kiss-and-run was inhibited by mutation of a homologous Ca^{2+} ligand in the C₂B domain of Syt IV. The Ca^{2+} sensitivity for full fusion was 5-fold higher with Syt I than Syt IV, but for kiss-and-run the Ca^{2+} sensitivities differed by a factor of only two. Syt thus regulates the choice between full fusion and kiss-and-run, with Ca^{2+} binding to the C₂A and C₂B domains playing an important role in this choice.

Amperometry resolves distinct kinetic steps in exocytosis. Full fusion proceeds through the opening of a fusion pore and the appearance of a prespike foot (PSF)^{7–9}, followed by a fusion pore dilation step to produce a spike in the amperometric record as the vesicle content is completely expelled (Fig. 1a). The dynamics of PSF can be altered in PC12 cells by transfection with either Syt I or Syt IV, suggesting that Syt has a regulatory interaction with the fusion pore⁹. In contrast to spikes, kiss-and-run events are revealed in amperometric recording as stand-alone feet (SAF), in which fusion pores open and close without dilating so that the event terminates by returning to baseline without a spike^{2,5} (Fig. 1a). PC12 cells transfected with Syt IV, an isoform of Syt that is induced by seizures^{10,11}, produce frequent SAF with amplitudes that are very small (~0.45 pA; Figs 1a and 2c, d) compared with PSF (~2.5 pA; Fig. 1a; see ref. 9). Similar SAF were seen in cells treated for 4–8 h with forskolin (50 μM) (Fig. 1a) to induce endogenous Syt IV expression¹⁰. SAF with similar amplitudes were seen in control and Syt I-transfected cells (Figs 1a and 2c, d), but the frequency was lower and the durations were shorter than in Syt IV-transfected and forskolin-treated cells (Fig. 1).

All fusion pore openings that maintain an amplitude of ~0.5 pA for a sufficient time to be detected (more than 5 ms) end by closing rather than dilating. Thus, these fusion pores seem to be dedicated exclusively to kiss-and-run. We are unable to ascertain whether fusion pore openings with amplitudes of PSF (2.5 pA) close in some instances without dilating. Events of this amplitude that are brief (~1 ms) would be difficult to distinguish from small spikes. Although we cannot state that 2.5-pA fusion pores are exclusively dedicated to full fusion, given the rare incidence of rectangular events of this amplitude, it is likely that most openings of 2.5-pA fusion pores lead to full fusion. SAF have a strikingly rectangular shape (Fig. 1a) that could be caused by a transient appearance of clusters of noradrenaline transporters in the plasma membrane.

However, the noradrenaline transporter blocker desipramine (10 μM), which blocked the loading of vesicles when incubated overnight with cells, failed to block SAF when applied for 15 min (data not shown). Furthermore, SAF were seen in permeabilized cells (see Fig. 5), so transporter contributions to SAF can be ruled out.

Spike and SAF frequencies varied from cell to cell. We used this variability to examine the relation between these two pathways. SAF and spike frequency were measured in many cells and plotted against one another. This plot showed a strong correlation in controls, in cells transfected with Syt I or Syt IV, and in cells treated with forskolin ($P < 0.0001$ for all) (Fig. 1b). Cells producing more spikes also produce more SAF, indicating that the two pathways employ, but do not compete for, common factors responsible for cell-to-cell variability. The slope of these plots is the frequency ratio and provides a quantitative index of the preference between these

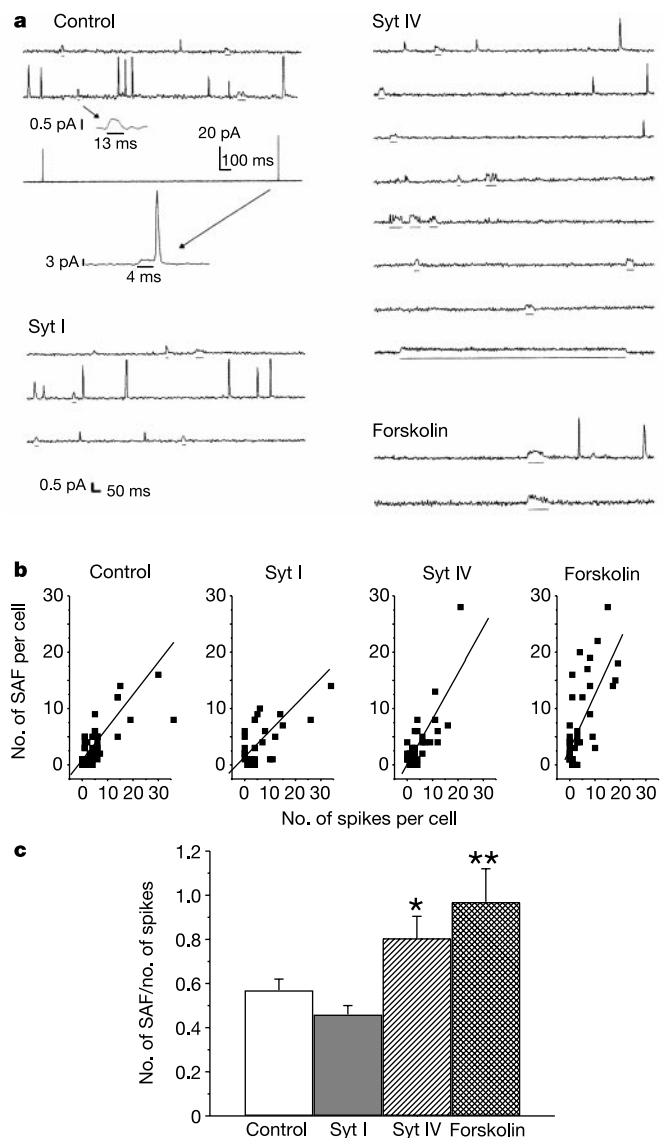


Figure 1 Spikes and SAF in PC12 cells. **a**, PC12 cells produced low-amplitude, long-duration SAF after KCl depolarization. By comparison, PSF are larger and briefer (expanded lower control trace). SAF in Syt IV-transfected and forskolin-treated cells were longer than SAF in control and Syt I-transfected cells (marked by horizontal lines below traces). **b**, Plots of SAF frequency against spike frequency, with each cell providing a single point. **c**, Ratio of SAF to spike frequency (slopes from **b**) (30–47 cells and 6–8 transfections for each, 3 batches of forskolin-treated cells; asterisk, $P < 0.05$; two asterisks, $P < 0.01$, relative to control and Syt I).

two pathways. This ratio for Syt IV (0.81 ± 0.09) was similar to that of forskolin-treated cells (0.97 ± 0.15), and both were higher than those for Syt I (0.46 ± 0.04) and control (0.57 ± 0.05) (Fig. 1c), indicating that Syt IV directs the exocytotic resources of a cell towards the kiss-and-run pathway.

PSF, with durations of ~ 1 – 2 ms (ref. 9), were much shorter than SAF (Figs 1a and 2a, b). SAF were 82 ± 11 ms in Syt IV-transfected cells and 40 ± 5 ms in forskolin-treated cells. SAF in control cells (20 ± 1 ms) and Syt I-transfected cells (21 ± 1 ms) (Figs 1a and 2a) were shorter, but still much longer than PSF. In contrast to durations, amplitudes remained similar at ~ 0.45 pA in transfected cells (Fig. 2c) and in forskolin-treated cells (data not shown). The distributions were approximately gaussian with similar means and variances (Fig. 2d). The distributions of SAF and PSF amplitudes had almost no overlap, supporting the view that these events arise from different populations of fusion pores. The distribution of SAF duration was exponential or multiexponential, as expected for the stochastic behaviour of single channels^{12,13} (Fig. 2b). Three exponentials were needed to fit the lifetime distribution in Syt IV-transfected cells, ($\tau = 21$ ms (71%), 120 ms (25%) and 1,250 ms (4%) (percentages of total events)) and forskolin-treated cells ($\tau = 8$ ms (80%), 28 ms (16%) and 228 ms (4%); $N = 296$). By contrast, single exponentials fitted the lifetime distribution of control ($\tau = 12$ ms) and Syt I-transfected cells ($\tau = 13$ ms). Thus, Syt IV alters SAF kinetics by allowing access to two additional conformational states. The single state seen in the absence of Syt IV might correspond to the briefest state seen in its presence.

The number of noradrenaline molecules in a vesicle in Syt IV-transfected cells undergoing full fusion can be estimated from the area of a spike as $76,100 \pm 2,500$ ($N = 1,074$) (ref. 9). The number of noradrenaline molecules escaping during a SAF is significantly larger, $117,000 \pm 17,000$ ($N = 419$; $P < 0.001$), but because SAF do not result in complete expulsion of a vesicle's content there must be considerably more noradrenaline in vesicles producing SAF. Analysis of immunolabelled dense-core vesicles revealed that the size of profiles positive for Syt IV was greater than of profiles

positive for Syt I (Fig. 3). These experiments indicated that both Syt I and Syt IV reside on dense-core vesicles. Cells transfected with Syt IV had many Syt IV-labelled vesicles, as well as Syt I-labelled vesicles at a lower density (Fig. 3a); some vesicles were labelled for both proteins (Fig. 3b). In Syt I-transfected cells more Syt I-labelled vesicles were seen, and Syt IV labelling was very rare. About 13% of both labels resided on dense-core vesicles. Of the remaining cytoplasmic label, some was probably on clear vesicles but the fraction could not be determined because vesicular membrane was not visible. The distribution of profile sizes labelled with Syt I and Syt IV overlapped but were not identical (Fig. 3c), with mean diameters of 56.5 ± 1.2 nm ($N = 142$) for Syt I and 63.9 ± 1.1 nm ($N = 162$; $P < 0.001$) for Syt IV (Fig. 3d) (implying a 45% difference in volume). These diameters are of the vesicle cores and are therefore smaller than the vesicle diameter. A size difference between vesicles containing Syt I and those containing Syt IV was reported previously¹⁴. Although quantitative comparison between ultrastructure and noradrenaline content is not possible, partly because of uncertainties in the noradrenaline content of SAF, these results are both consistent with the hypothesis that kiss-and-run and full fusion arise from vesicles with different but overlapping size distributions.

The Ca^{2+} -dependent phospholipid binding exhibited by Syt I cannot be detected in Syt IV (refs 15–17). *Drosophila* Syt I and Syt IV can both bind syntaxin in a Ca^{2+} -promoted manner, although the activity of Syt IV was somewhat lower¹⁷. It is well established that Ca^{2+} promotes the binding of rat Syt I synaptosome-associated protein of relative molecular mass 25K (SNAP-25) and syntaxin¹⁸. Here we assessed the interactions of rat Syt IV, and observed binding to isolated SNAP-25 and also to SNAP-25/syntaxin heterodimers; this interaction was weakly promoted by Ca^{2+} (Supplementary Information). A key difference between Syt I and Syt IV is the substitution of serine for an aspartate residue (Asp 230) that functions as the third acidic Ca^{2+} ligand in the C_2A domain of Syt I (ref. 15). This residue coordinates with two Ca^{2+} ions¹⁹ and its absence probably accounts for the failure of the C_2A domain of Syt IV to bind lipid. By contrast, the second Ca^{2+} -binding module of

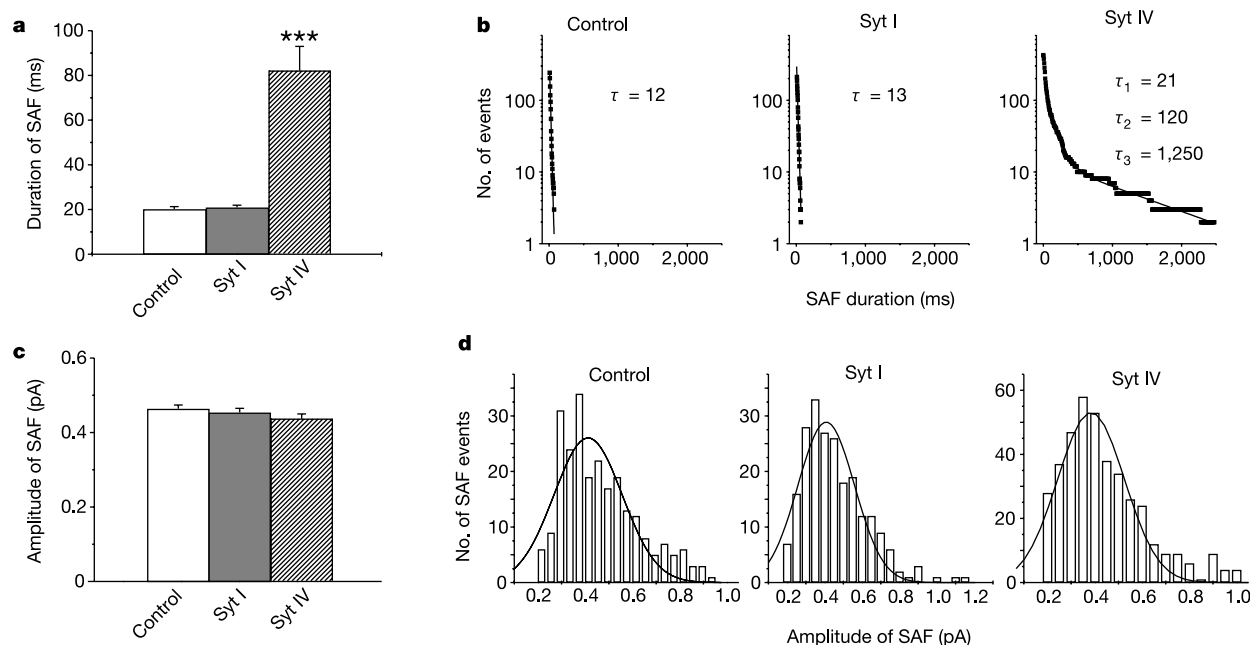


Figure 2 SAF properties. **a**, **b**, Mean duration (**a**) and lifetime distribution (**b**) of SAF for control cells and for cells transfected with Syt I and Syt IV. **c**, **d**, Mean SAF amplitudes (**c**) and amplitude distributions (**d**) for control, Syt I and Syt IV with gaussian fits ($\chi^2 \sim 18$ for

all) (228–427 SAF, numbers of cells and transfections as in Fig. 1; three asterisks, $P < 0.001$).

Syt, the C₂B domain, has conserved Ca²⁺ ligands in both Syt I and Syt IV (refs 10, 15). The role of Ca²⁺ binding was tested by making a D230S mutation in Syt I and the reverse mutation at the corresponding site in Syt IV (S244D). Mutations were also tested in the corresponding locations of C₂B in Syt I (D363N) and Syt IV (D378N). SAF in cells transfected with Syt I-D230S (Fig. 4a) strongly resembled SAF in cells transfected with wild-type Syt IV. There was no change in amplitude, but the relative frequency of SAF to spikes rose to 0.97 ± 0.07 (Fig. 4b), a value indistinguishable from 0.81 ± 0.09 obtained for wild-type Syt IV (Fig. 1b, c). The D230S mutation also increased the SAF duration to 84 ± 16 ms (Fig. 4c), which is indistinguishable from the value of 82 ± 11 ms with wild-type Syt IV (Fig. 2a). The SAF lifetime distribution was no longer single exponential, as in Syt I, but was a sum of three exponentials (Fig. 4d), again like wild-type Syt IV (Fig. 2b). The opposite mutation in Syt IV (S244D) also left the amplitudes unchanged and partly converted SAF to those seen with Syt I (Fig. 4a). The frequency ratio was similar to that with Syt IV, but the duration was reduced to 34 ± 9 ms (Fig. 4e, f). The lifetime distribution was reduced from a sum of three to a sum of two exponentials (Fig. 4g).

Mutating the Ca²⁺ ligand in C₂B has the opposite effect in Syt IV, reducing the frequency ratio (Fig. 4e), shortening SAF duration (Fig. 4f) and changing the lifetime distribution to a single exponential (Fig. 4g). Thus, SAF seen with Syt IV-D378N are almost identical to those seen with wild-type Syt I. The homologous mutation in Syt I reduced SAF duration but had almost no effect on frequency ratio (Fig. 4b–d). These results from mutant proteins indicate that Ca²⁺ binding to the C2 domains has a key role in the choice between full fusion and kiss-and-run.

Syt is widely viewed as an important transducer of Ca²⁺ signals in regulated exocytosis^{18,20}. The results on SAF so far indicated an important role for Ca²⁺ binding to Syt, so we examined the Ca²⁺ dependence of SAF and spikes. In permeabilized cells the direct application of Ca²⁺ triggered exocytosis, eliciting both spikes and SAF (Fig. 5a). Spike and SAF frequency were separately measured as the slope of the cumulative event count against time⁹, and plotted

against [Ca²⁺] (Fig. 5b). Results for Syt I and controls were similar (Fig. 5b). Plots were well fitted by the Hill equation, providing parameters for apparent affinity and cooperativity of the Ca²⁺ sensor for each mode of exocytosis. For spikes, Syt I had a significantly higher apparent Ca²⁺ affinity than Syt IV ($EC_{50} = 0.65 \pm 0.05 \mu\text{M}$ for Syt I and $3.4 \pm 1.0 \mu\text{M}$ for Syt IV; $P < 0.01$), which is consistent with the idea that Syt IV is a poorer Ca²⁺ sensor than Syt I. Furthermore, for [Ca²⁺] $\geq 1 \mu\text{M}$ the ratio of SAF frequency to spike frequency was higher for Syt IV, consistent with Fig. 1b, c. The shift in spike sensitivity supports the idea that Ca²⁺ binding to C₂A triggers full fusion. Both Syt I and Syt IV had low Hill slopes, indicating noncooperative action ($N = 0.78 \pm 0.06$ and 0.80 ± 0.10 for Syt I and Syt IV, respectively). Strikingly, the maximal spike frequencies at high [Ca²⁺] were comparable (0.38 ± 0.01 and 0.36 ± 0.03 spikes s⁻¹ for Syt I and Syt IV, respectively; $P = 0.58$). Thus, Syt IV functions as well as Syt I when its Ca²⁺ binding deficit is compensated for by elevated [Ca²⁺]. This is consistent with the

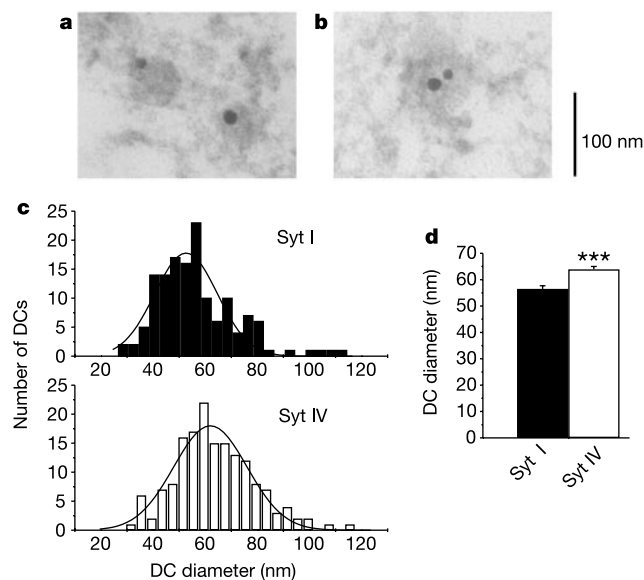


Figure 3 Immunogold electron microscopy showed dense-core vesicle localization of Syt I (20 nm) and Syt IV (10 nm). **a**, In Syt IV-transfected cells, one vesicle is labelled with Syt I and another with Syt IV. **b**, A single vesicle is labelled with Syt I and Syt IV. **c**, Distributions of Syt I- and Syt IV-labelled dense-core (DC) diameter with best fitting gaussians ($\chi^2 \sim 3.2$ for both). **d**, The mean dense-core diameters for Syt I and Syt IV from the gaussian fits in **c** (three asterisks, $P < 0.001$).

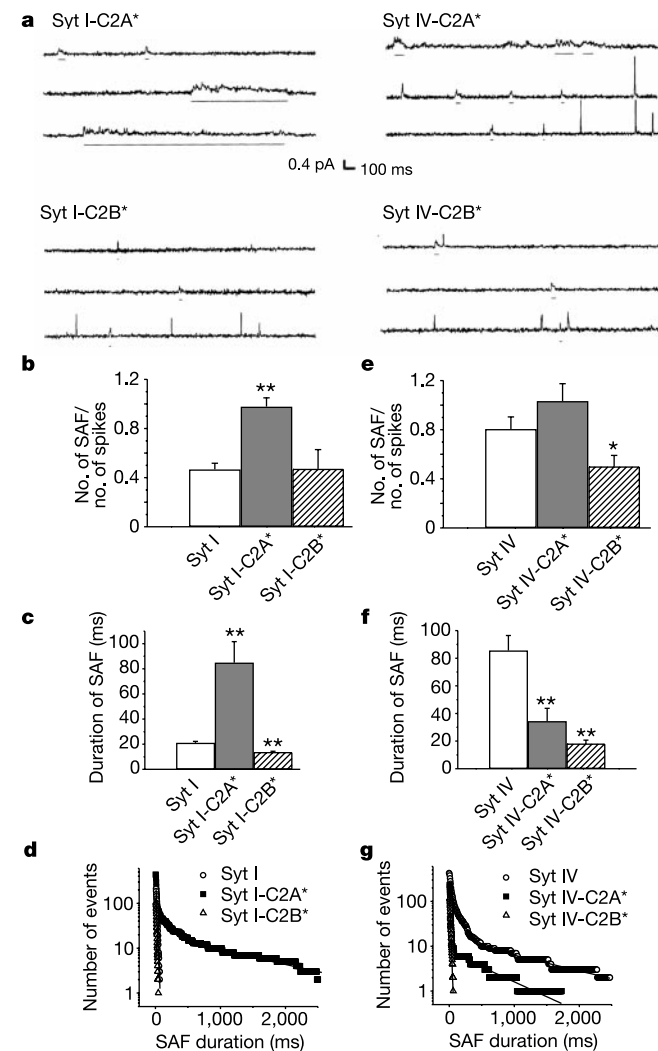


Figure 4 Ca²⁺ ligand mutations in C₂A and C₂B. **a**, SAF had very different durations with Syt I-D230S (Syt I-C₂A*), Syt I-D363N (Syt I-C₂B*), Syt IV-S244D (Syt IV-C₂A*) and Syt IV-D378N (Syt IV-C₂B*) (marked with horizontal lines). **b–g**, The frequency ratio (**b**, **e**), mean duration (**c**, **f**) and lifetime distribution (**d**, **g**) of SAF for the indicated proteins. For Syt I-C₂A* $\tau_1 = 11$ ms, $\tau_2 = 190$ ms, $\tau_3 = 1,366$ ms; for Syt I-C₂B* $\tau = 9$ ms; for Syt IV-C₂A* $\tau_1 = 9$ ms, $\tau_2 = 655$ ms; for Syt IV-C₂B* $\tau = 11$ ms (104–440 events, 21–42 cells and 6–8 transfections; asterisk, $P < 0.05$; two asterisks, $P < 0.01$, relative to wild type).

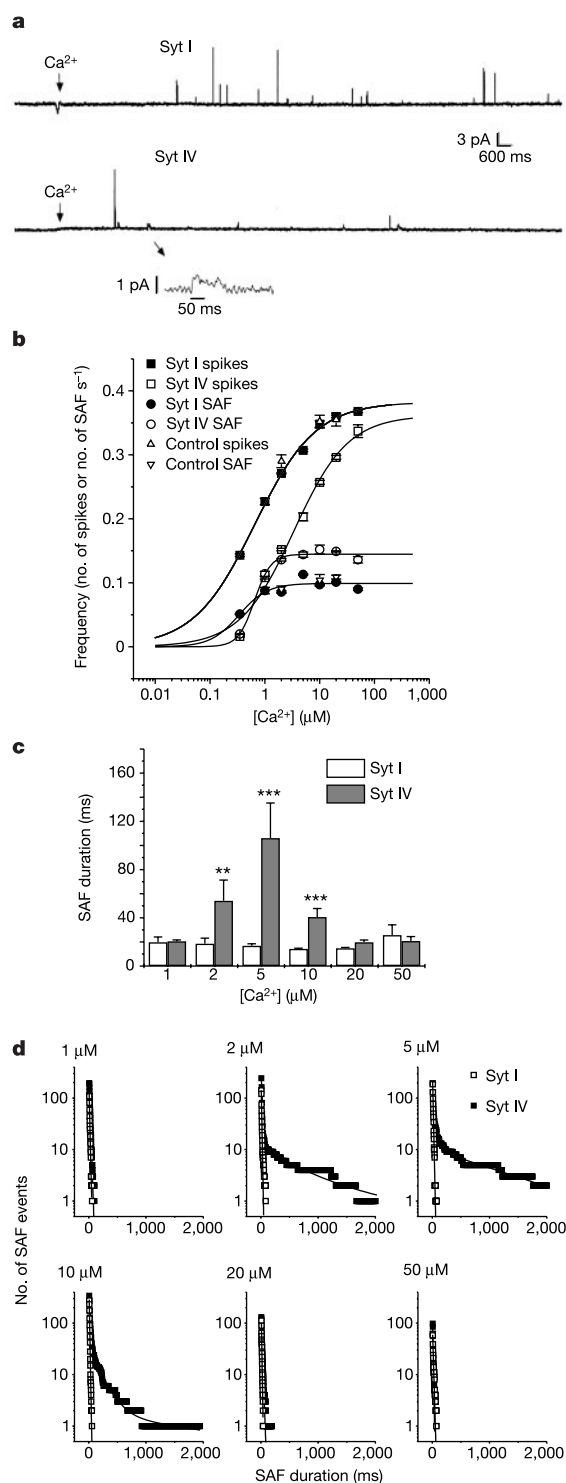


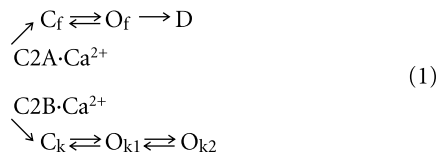
Figure 5 Secretion from permeabilized cells. **a**, Spikes and SAF evoked by Ca^{2+} (5 μM , applied for 22 s starting at the arrow) in permeabilized cells transfected with Syt I and Syt IV. The expanded trace below shows a SAF in a Syt IV-transfected cell. **b**, Ca^{2+} dependence of spike and SAF frequency in cells transfected with Syt I and Syt IV, fitted to the Hill equation (χ^2 0.00032; fit includes zero). **c**, Mean SAF duration versus $[\text{Ca}^{2+}]$ (two asterisks, $P < 0.01$; three asterisks, $P < 0.001$, for Syt I versus Syt IV). **d**, Lifetime distributions of SAF, fitted with single exponentials for Syt I ($\tau = 7$ –11 ms) and Syt IV in 1, 20 and 50 μM Ca^{2+} ($\tau = 7$ –14 ms). Three exponential fits in 2, 5 and 10 μM Ca^{2+} gave $\tau_1 = 9$ –14 ms, $\tau_2 = 71$ –484 ms, $\tau_3 = 1,200$ –2,484 ms (61–275 events, 12–31 cells and 3–7 transfections).

finding that Syt IV can rescue some degree of synaptic transmission in *Drosophila* lacking Syt I (ref. 21).

SAF were activated by submicromolar Ca^{2+} and had a higher Ca^{2+} sensitivity than spikes for both Syt I ($\text{EC}_{50} = 0.33 \pm 0.07 \mu\text{M}$; $P < 0.001$ compared with spikes) and Syt IV ($\text{EC}_{50} = 0.65 \pm 0.04 \mu\text{M}$; $P < 0.01$ compared with spikes). The Ca^{2+} sensitivities for SAF differ by only a factor of two between the two Syt isoforms, compared with a factor of 5 noted above for spikes. Moreover, the Ca^{2+} dependence for SAF showed positive cooperativity, with Hill slopes of 1.7 ± 0.7 and 2.9 ± 0.4 for Syt I and Syt IV, respectively. Thus, full fusion and kiss-and-run are both Ca^{2+} -dependent forms of exocytosis, but Ca^{2+} transduction for these two processes is clearly different. SAF amplitudes were the same for $[\text{Ca}^{2+}]$ from 0.35 to 50 μM for both Syt I and Syt IV. $[\text{Ca}^{2+}]$ had no significant effect on SAF duration (Fig. 5c) and closing kinetics (Fig. 5d, open squares) in Syt I ($P = 0.11$ by one-way analysis of variance (ANOVA)), but in Syt IV the SAF duration showed a strong bell-shaped $[\text{Ca}^{2+}]$ dependence, with a maximum at 5 μM (Fig. 5c; $P = 0.0003$ by one-way ANOVA). In addition, the SAF lifetime distribution was multiexponential only at intermediate $[\text{Ca}^{2+}]$ (Fig. 5d). Comparison of these results with those in Fig. 2 allowed us to estimate $[\text{Ca}^{2+}]$ near the Ca^{2+} sensor as 2–10 μM in intact cells during KCl depolarization.

These findings clarify the role of Syt in two pathways of regulated exocytosis. Syt I favours full fusion and Syt IV favours kiss-and-run. The fusion pores associated with these pathways differ markedly in their size and stability: full fusion is mediated by large pores with brief openings, and kiss-and-run is mediated by small pores with long openings. The failure of kiss-and-run fusion pores to dilate suggests that these structures lack the capacity to incorporate lipids necessary for pore expansion. By contrast, full-fusion pores can incorporate lipids very rapidly and reliably. The two pathways also differ in terms of Ca^{2+} sensitivity. Kiss-and-run is triggered by lower $[\text{Ca}^{2+}]$ with greater cooperativity, indicating fundamental differences in the Ca^{2+} -sensing mechanisms of these two forms of transmitter release.

Kiss-and-run is enhanced by weakened Ca^{2+} binding to C_2A of Syt I and inhibited by homologous mutations in C_2B of Syt IV. The high Ca^{2+} sensitivity of kiss-and-run allows it to occur even when C_2A domains bind Ca^{2+} effectively, as indicated by the significant numbers of SAF in control cells and Syt I-transfected cells. The distinct functions of the C_2A and C_2B domains in the two exocytotic pathways are illustrated in Equation (1).



Full fusion (upper pathway, subscript f) is favoured by Ca^{2+} binding to C_2A . Kiss-and-run (lower pathway, subscript k) is favoured by Ca^{2+} binding to C_2B . The closed states (C) form in a step from the left. Closed states can go to open states (O), of which only O_f can dilate (D). On the basis of the appearance of three exponentials in the lifetime distribution and closures within bursts we tentatively designate SAF as being composed of three states, one closed and two open. Because the C_2B domain binds only two Ca^{2+} ions and the Hill slope for Ca^{2+} triggered kiss-and-run was greater than 2 for Syt IV, this suggests either that Syt triggers kiss-and-run as an oligomer or that there is a weak contribution made by the residual Ca^{2+} binding of C_2A (ref. 22). Although this model does not account for the high SAF frequencies for Syt IV-S244D and Syt I-D363N, Syt I and Syt IV are only 40% homologous¹¹, so it is likely that other parts of the protein also contribute to the choice between

full fusion and kiss-and-run.

Syt IV mRNA has been reported in the hippocampus¹⁰, and this might be relevant to reports of kiss-and-run in hippocampal cultures^{23,24}. The long duration of the kiss-and-run events will allow neurotransmitter to escape, but the small apparent pore size will make that release very slow. If Syt IV functions similarly at synapses when expressed in neurons, the slow release will elevate ambient extracellular neurotransmitter concentrations to desensitize postsynaptic receptors without generating synaptic responses (discussed in ref. 3). Thus, the presence of Syt IV in a nerve terminal could render the synapse silent. By desensitizing postsynaptic receptors, the slow kiss-and-run release resulting from the upregulation of Syt IV during seizures^{10,11} will reduce synaptic responses and cause a global depression of activity to counteract hyperexcitability. □

Methods

Molecular biology

Syt I-D230S, Syt I-D363N, Syt IV-S244D and Syt IV-D378N were prepared by modified sequential polymerase chain reaction (PCR) (ref. 25), using pIRES2EGFP (Clontech) encoding wild-type protein as template. The terminal primers annealed the upstream *NheI* site and the downstream *BamHI* site for Syt I mutants and the upstream *MluI* site and the downstream *BglII* site for Syt IV. These sites were used to subclone the PCR product back into pIRES2EGFP. Mutants were confirmed by sequencing.

Cell culture

PC12 cells were cultured and loaded with noradrenaline as described previously⁹. For forskolin-treated cells (50 μ M for 4–8 h) and freeze–thaw–permeabilized cells, Petri dishes were precoated with 50 μ g ml^{−1} collagen I and 50 μ g ml^{−1} poly-d-lysine (Becton Dickinson). SAF were indistinguishable between precoated and non-precoated dishes, and between control and non-transfected cells.

Transfection

Cells were transfected with 50 μ g DNA as described previously⁹ or with an ECM 830 electroporator (BTX Inc.) with 230-V, 5-ms pulses. Release was examined 60–96 h after transfection. Control cells were transfected with pIRES2EGFP lacking an insert at the cytomegalovirus promoter.

Cell permeabilization

Cells were permeabilized by freezing and thawing²⁶. After being washed three times with (in mM) 120 potassium glutamate, 20 potassium acetate, 2 EGTA, 20 HEPES pH 7.2, cells were bathed in this solution. A dish wrapped in Parafilm was immersed in liquid nitrogen for 10 s, thawed at 37 °C for 30 min, rinsed, and bathed in the above solution but with 121 mM potassium glutamate and 0.2 mM EGTA. GFP fluorescence, still visible after permeabilization, was used to select transfected cells. Secretion could be detected for up to 2 h.

Amperometry

Amperometric recording and KCl depolarization in intact cells follow previously described methods^{9,27}. In permeabilized cells, Ca²⁺ solutions were applied with a Picospritzer (General Valve, Fairfield, New Jersey) to trigger exocytosis. Potassium glutamate buffer (pH 7.2) with 0.2 mM EGTA was prepared with the desired free [Ca²⁺] as measured with an electrometer and Ca²⁺ electrode (Orion).

Immunoelectron microscopy

Cells transfected with Syt I and Syt IV were fixed for 2 h with 4% paraformaldehyde, 0.1% glutaraldehyde in 0.1 M phosphate buffer pH 7.4, rinsed three times with this buffer, and dehydrated in a graded ethanol series to 100%. Samples were infiltrated with LR White and embedded in gelatin. The embedding medium was polymerized for 48 h at 60 °C.

Thin (60–90-nm) sections were blocked in phosphate buffer, 0.1% Triton X-100, 0.1% fish gelatin and 0.1% BSA, exposed to the primary antibody diluted in blocking buffer (1:100) for 12 h at 4 °C, rinsed five times in blocking buffer, and exposed to the secondary antibody (either 10-nm gold-conjugated goat anti-rabbit for the rabbit polyclonal antibody against Syt IV (details available from the authors on request) or 20 nm gold-conjugated goat anti-mouse for the mouse monoclonal antibody against Syt I (ref. 28)) at a dilution of 50:1 in blocking buffer. The samples were rinsed five times each in blocking buffer and distilled water. Control samples stained with the secondary antibody showed no gold labelling.

For electron microscopy (with a Philips CM 120 electron microscope), some samples were stained with routine concentrations of uranyl acetate and lead citrate, and others were left unstained for the better visualization of gold particles. Labelled vesicles were photographed at random for analysis before estimation of particle size. Because dense cores were clearly visible and membranes were poorly preserved, size measurements were made of the dense core.

Data analysis

Spikes were analysed as described previously⁹. Events of 2 pA or more were counted as spikes. To view SAF, records were filtered at 100 Hz (eight-pole Bessel) to reduce noise. An in-house computer program was used to detect and analyse SAF by identifying events greater than four times root-mean-square noise and less than 2 pA. For long SAF, a

rectangular shape could be recognized and distinguished from a spike arising from a distant release site. The start and end points were taken as the times at which signals departed from and returned to $\times 1$ root mean square above baseline. The duration and area were taken as the time interval and integral between these two points, respectively. SAF amplitude was computed by dividing area by duration. SAF of less than 0.2 pA were excluded. Cutoff duration for SAF analysis was 7.5 ms for measuring amplitudes and durations, and 5 ms to count SAF for frequencies. In the bursting of longer SAF (for example with Syt IV), an event was counted as one SAF if the intra-burst closed times were less than 50 ms.

SAF characteristics are reported as means $\pm$ s.e. Unpaired *t*-test (for comparing two groups) or one-way ANOVA with the Newman–Keuls multiple comparison test (for comparing more than two groups) were used to evaluate statistical significance. Because the SAF duration distribution was not gaussian, data were transformed logarithmically for statistical analysis.

Received 9 March; accepted 18 June 2003; doi:10.1038/nature01857.

1. Neher, E. & Marty, A. Discrete changes of cell membrane capacitance observed under conditions of enhanced secretion in bovine adrenal chromaffin cells. *Proc. Natl Acad. Sci. USA* **79**, 6712–6716 (1982).
2. Alvarez de Toledo, G., Fernandez-Chacon, R. & Fernandez, J. M. Release of secretory products during transient vesicle fusion. *Nature* **363**, 554–558 (1993).
3. Klyachko, V. A. & Jackson, M. B. Capacitance steps and fusion pores of small and large-dense-core vesicles in nerve terminals. *Nature* **418**, 89–92 (2002).
4. Schneider, S. W. Kiss and run mechanism in exocytosis. *J. Membr. Biol.* **181**, 67–76 (2001).
5. Albillos, A. *et al.* The exocytotic event in chromaffin cells revealed by patch amperometry. *Nature* **389**, 509–512 (1997).
6. Ales, E. *et al.* High calcium concentrations shift the mode of exocytosis to the kiss-and-run mechanism. *Nature Cell Biol.* **1**, 40–44 (1999).
7. Chow, R. H., von Rüden, L. & Neher, E. Delay in vesicle fusion revealed by electrochemical monitoring of single secretory events in adrenal chromaffin cells. *Nature* **356**, 60–63 (1992).
8. Schroeder, T. J. *et al.* Temporally resolved, independent stages of individual exocytotic secretion events. *Biophys. J.* **70**, 1061–1068 (1996).
9. Wang, C.-T. *et al.* Synaptotagmin modulation of fusion pore kinetics in regulated exocytosis of dense-core vesicles. *Science* **294**, 1111–1115 (2001).
10. Vician, L. *et al.* Synaptotagmin IV is an immediate early gene induced by depolarization in PC12 cells and in brain. *Proc. Natl Acad. Sci. USA* **92**, 2164–2168 (1995).
11. Ferguson, G. D., Vician, L. & Herschman, H. R. Synaptotagmin IV: Biochemistry, genetics, behavior, and possible links to human psychiatric disease. *Mol. Neurobiol.* **23**, 173–185 (2001).
12. Jackson, M. B. Stationary single channel analysis. *Methods Enzymol.* **207**, 729–746 (1992).
13. Colquhoun, D. & Hawkes, A. G. On the stochastic properties of bursts of single ion channel openings and of clusters of bursts. *Phil. Trans. R. Soc. Lond. B* **300**, 1–59 (1982).
14. Eaton, B. A., Haugwitz, M., Lau, D. & Moore, H.-P. H. Biogenesis of regulated exocytotic carriers in neuroendocrine cells. *J. Neurosci.* **20**, 7334–7344 (2000).
15. von Poser, C., Ichtchenko, K., Shao, X., Rizo, J. & Südhof, T. The evolutionary pressure to inactivate: A subclass of synaptotagmins with an amino acid substitution that abolishes Ca²⁺ binding. *J. Biol. Chem.* **272**, 14314–14319 (1997).
16. Bai, J., Wang, P. & Chapman, E. R. C2A activates a cryptic Ca²⁺-triggered membrane penetration activity within the C2B domain of synaptotagmin I. *Proc. Natl Acad. Sci. USA* **99**, 1665–1670 (2002).
17. Littleton, J. T., Serano, T. L., Rubin, G. M., Ganetzky, B. & Chapman, E. R. Synaptic function modulated by changes in the ratio of synaptotagmin I and IV. *Nature* **400**, 757–760 (1999).
18. Chapman, E. R. Synaptotagmin: A Ca²⁺ sensor that triggers exocytosis? *Nature Rev. Mol. Cell Biol.* **3**, 498–508 (2002).
19. Ubach, J., Zhang, X., Shao, X., Südhof, T. C. & Rizo, J. Ca²⁺ binding to synaptotagmin: How many Ca²⁺ ions bind to the tip of a C2-domain? *EMBO J.* **17**, 3921–3930 (1998).
20. Fernandez-Chacon, R. *et al.* Synaptotagmin I functions as a calcium regulator of release probability. *Nature* **410**, 41–49 (2001).
21. Robinson, I. M., Ranjan, R. & Schwartz, T. L. Synaptotagmins I and IV promote transmitter release independently of Ca²⁺ in the C₂A domain. *Nature* **418**, 336–340 (2002).
22. Davis, A. F. *et al.* Kinetics of synaptotagmin responses to Ca²⁺ and assembly with the core SNARE complex onto membranes. *Neuron* **24**, 363–376 (1999).
23. Klingauf, J., Kavalali, E. T. & Tsien, R. W. Kinetics and regulation of fast endocytosis at hippocampal synapses. *Nature* **394**, 581–585 (1998).
24. Stevens, C. F. & Williams, J. H. 'Kiss and run' exocytosis at hippocampal synapses. *Proc. Natl Acad. Sci. USA* **97**, 12828–12833 (2000).
25. Cormack, B. in *Current Protocols in Molecular Biology* (eds Ausubel, F. M. *et al.*) 8.5.7–8.5.9 (Wiley, New York, 1994).
26. Klensch, V. A., Kowalchuk, J. A. & Martin, T. F. J. Large dense-core vesicle exocytosis in PC12 cells. *Methods* **16**, 204–208 (1998).
27. Chow, R. H. & von Rüden, L. in *Single-Channel Recording* (eds Sakmann, B. & Neher, E.) 245–275 (Plenum, New York, 1995).
28. Brose, N., Petrenko, A. G., Südhof, T. C. & Jahn, R. Synaptotagmin: A calcium sensor on the synaptic vesicle surface. *Science* **256**, 1021–1025 (1992).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank L. A. Schuler for support and access to facilities; R. Jahn and S. Engers for antibodies against Syt I; R. Grishanin, C. A. Earles, J. A. Kowalchuk, and X. Han for discussions; and B. August of the University of Wisconsin Electron Microscopy Facility for performing electron microscopy. This work was supported by grants from the NIH to E.R.C., T.F.J.M. and M.B.J., E.R.C. is a Pew Scholar.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.B.J. (mjackson@physiology.wisc.edu).

Generation of prion transmission barriers by mutational control of amyloid conformations

Peter Chien^{1,2}, Angela H. DePace², Sean R. Collins²
& Jonathan S. Weissman^{1,2}

¹Graduate Group in Biophysics, ²Howard Hughes Medical Institute, Department of Cellular and Molecular Pharmacology, and Department of Biochemistry and Biophysics, University of California–San Francisco, San Francisco, California 94143-2240, USA

Self-propagating β -sheet-rich protein aggregates are implicated in a wide range of protein-misfolding phenomena, including amyloid diseases and prion-based inheritance¹. Two properties have emerged as common features of amyloids. Amyloid formation is ubiquitous: many unrelated proteins form such aggregates and even a single polypeptide can misfold into multiple forms^{2–6} — a process that is thought to underlie prion strain variation⁷. Despite this promiscuity, amyloid propagation can be highly sequence specific: amyloid fibres often fail to catalyse the aggregation of other amyloidogenic proteins^{8,9}. In prions, this specificity leads to barriers that limit transmission between species^{7,8,10–12}. Using the yeast prion $[PSI^+]$ ¹³, we show *in vitro* that point mutations in Sup35p, the protein determinant of $[PSI^+]$, alter the range of ‘infectious’ conformations, which in

turn changes amyloid seeding specificity. We generate a new transmission barrier *in vivo* by using these mutations to specifically disfavour subsets of prion strains. The ability of mutations to alter the conformations of amyloid states without preventing amyloid formation altogether provides a general mechanism for the generation of prion transmission barriers and may help to explain how mutations alter toxicity in conformational diseases.

The yeast prion $[PSI^+]$ ¹³ is a non-mendelian element caused by self-propagating aggregates of the translation-termination factor Sup35p, which leads to a nonsense suppression phenotype. In strains containing an *ade1* gene with a nonsense mutation, $[PSI^+]$ colonies are white and grow on media lacking adenine, whereas $[psi^-]$ colonies are red and require adenine. $[PSI^+]$ can be induced *de novo* by transient overexpression of the modular amino-terminal

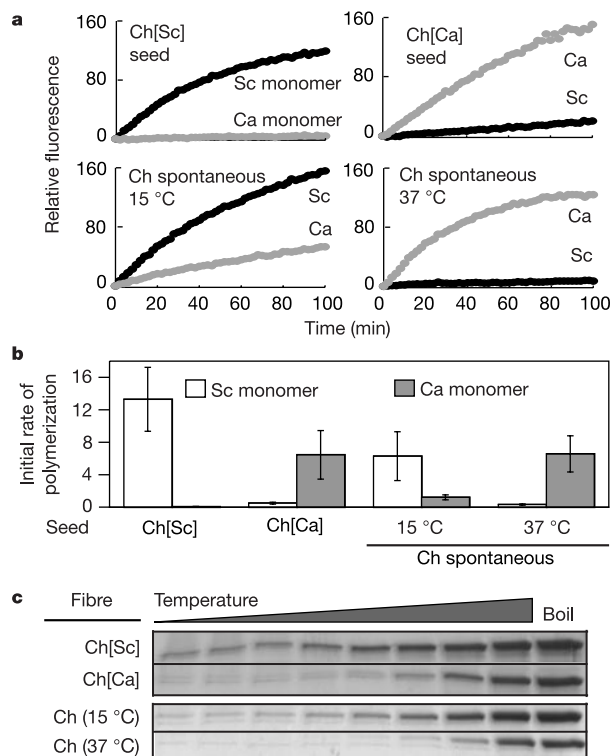


Figure 1 Conformation of chimaeric prion fibres is sensitive to polymerization conditions. **a**, Amyloid polymerization curves monitored by thioflavin-T fluorescence. Ch fibres were used to seed Sc or Ca monomer as labelled. The Ch fibres were generated either by spontaneous polymerization at the indicated temperature or by templating with an Sc or Ca seed, as indicated in brackets. **b**, Initial rates of polymerization from data in **a** (at least two independent measurements). **c**, Thermal stability of the indicated Ch fibre types as determined by SDS-PAGE (see Methods and Supplementary Fig. S3). Total recoverable protein level was determined by boiling samples (final lane).

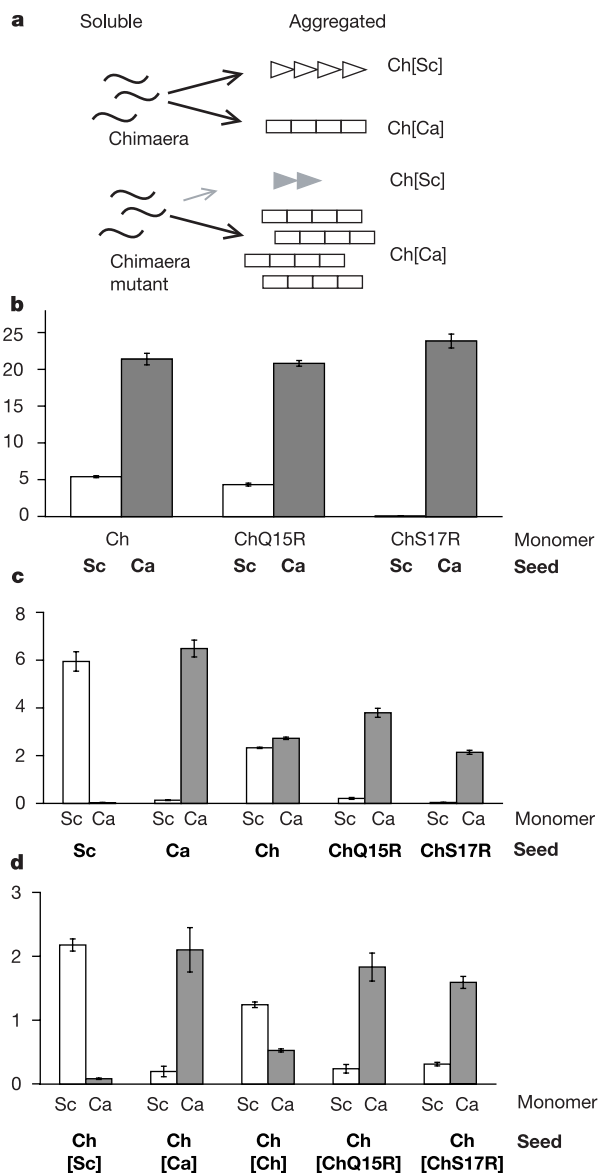


Figure 2 Mutations create a transmission barrier *in vitro*. **a**, Schematic of model wherein mutations favour formation of the Ch[Ca] state by specifically slowing the Ch[Sc] pathway. **b–d**, Specificity of seed as determined by initial rates of polymerization of the indicated monomer. In **b** and **c**, seeds were formed by spontaneous polymerization (at 25 °C) of the indicated Sup35 prion domain. In **d**, seeds were produced by templating Ch with the fibres indicated in brackets. The difference in specificity between Ch (**c**) and Ch[Ch] (**d**) is probably due to differential amplification of the mixed population of seeds generated during Ch polymerization.

glutamine/asparagine-rich prion domain, which is necessary for prion propagation and forms self-seeding amyloid fibres *in vitro*^{4,14}. Although the prion domain of Sup35 is conserved across a range of budding yeast, $[PSI^+]$ 'infectivity' is often limited by transmission barriers^{8,10–12} analogous to the species barriers seen in mammalian prions. For example, prion domains of Sup35p from *Saccharomyces cerevisiae* (Sc) and *Candida albicans* (Ca), although independently capable of forming heritable conformations, do not cross-seed each other *in vivo* or *in vitro*⁸. (Prion domains are referred to by two-letter species abbreviations and mutations, where appropriate; see Supplementary Table S1.) $[PSI^+]$, like mammalian prions and other yeast prions, exhibits a range of heritable phenotypic strain variants¹⁵, which are linked to differences in the conformation of the infectious prion protein, although cellular factors might also contribute to their propagation^{7,13}. $[PSI^+]$ strain variants differ in mitotic stability¹⁵, in interactions with the cellular chaperone machinery¹⁶, and in solubility and activity of the Sup35p protein^{15,17–19}. Sup35p aggregates purified from different $[PSI^+]$ strains also differ in their ability to seed purified Sc polymerization *in vitro*¹⁹. In addition, strains can have an important role in determining the specificity of prion transmission: $[PSI^+]$ prion strains differ greatly in their ability to recruit Sup35 mutants²⁰. Similarly, a chimaeric prion (Ch) that comprises the first forty amino acids of Sc fused to the remainder of the Ca prion domain is able to form at least two amyloid conformations, one that seeds Sc (termed Ch[Sc])

and another that seeds Ca (Ch[Ca])²¹. (Different amyloid conformations are referred to by the fibre protein followed by the seed used to initiate polymerization in brackets; see Supplementary Table S1.) This link between prion strains and transmission barriers is likely to be general, as the mammalian prion species barrier is also strain dependent⁷.

Although Ch polymerization induced by an Sc or Ca template robustly produces the Ch[Sc] or Ch[Ca] conformations²¹ (Fig. 1), we show here that conformations of Ch fibres formed in the absence of seed are highly sensitive to the polymerization conditions. As assessed by both seeding specificity (Fig. 1a, b) and by more direct physical measurements, such as protease protection (Supplementary Fig. S1) or resistance to denaturation (Fig. 1c), polymerization of Ch at 15 °C shows a strong bias for forming the Ch[Sc] conformation, whereas Ch seeds formed at 37 °C adopt a Ch[Ca] conformation. Polymerization at intermediate temperatures yields a mixture of both conformations (see below). Once formed, the two Ch conformations are highly stable and propagate robustly: for example, Ch that polymerized at 15 °C with seeds formed at 37 °C retained the seeding specificity of Ch polymerized at 37 °C (Supplementary Fig. S2).

If, as suggested by the above data, the choice between different prion forms is under kinetic control, then mutations that differentially affect the rate of formation of individual amyloid conformations should shift the spectrum of self-propagating prion states.

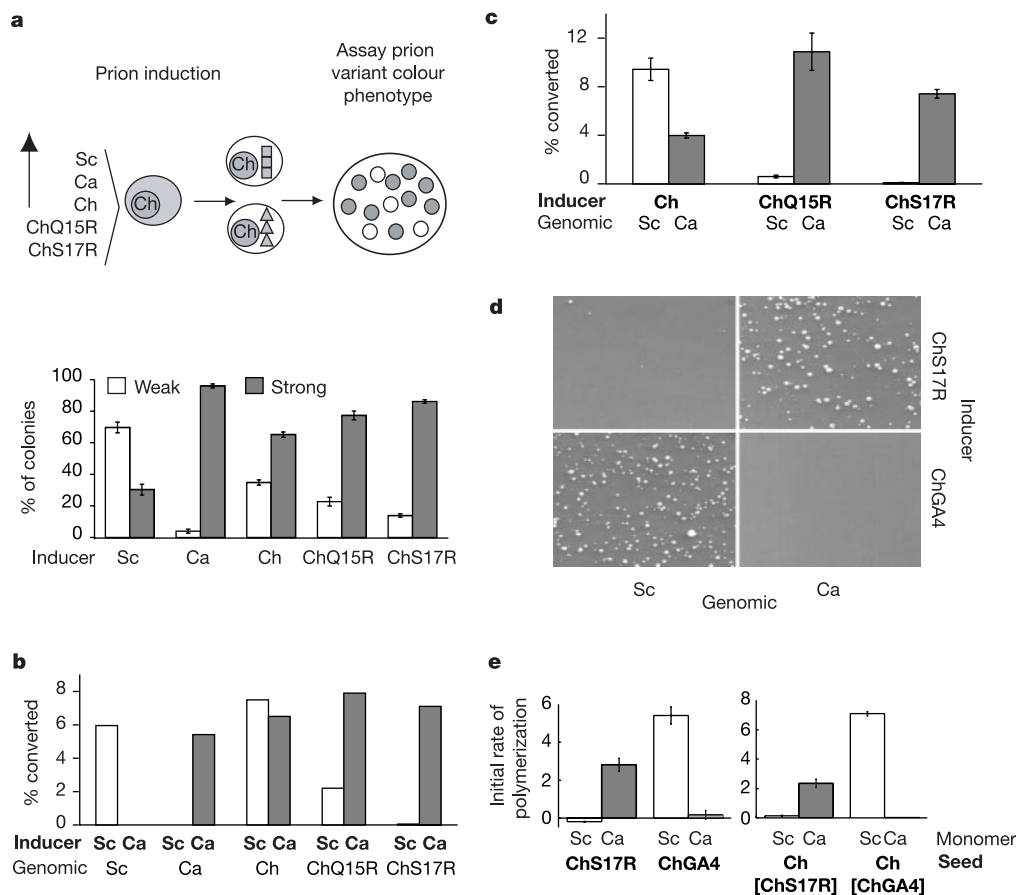


Figure 3 Mutations create a transmission barrier *in vivo* by shifting strain preference of chimaeric prion. **a**, Prion formation was induced in yeast expressing Ch Sup35 by transient overexpression of the indicated prion domain (top schematic). The relative amounts of 'strong' (grey) and 'weak' (white) strains are shown below. **b–d**, Relative efficiency of prion induction following overexpression of the indicated prion domains in

yeast expressing the indicated genomic Sup35 as assayed by growth on minus ADE media (see Methods). **e**, Initial rates of polymerization of indicated Sup35 prion domain with indicated seeds produced by either spontaneous polymerization of ChS17R or ChGA4 at 25 °C (left), or by Ch fibres generated by templating with seeds indicated in brackets (right).

In the case of the Ch protein, where the Ch[Sc] and Ch[Ca] prion forms have different seeding specificity, mutations that selectively slow one pathway should generate a transmission barrier (Fig. 2a). We tested this by examining whether a series of point mutations known to slow Sc prion formation²² would also specifically inhibit the Ch[Sc] pathway. We grafted two of these mutations onto the Sc-derived N terminus of the Ch prion: Q15R, which moderately disables Sc prion formation (yielding ChQ15R), and S17R, which has a stronger effect^{20,22} (yielding ChS17R).

As predicted, these mutations create a transmission barrier against wild-type Sc protein *in vitro* by disfavoured the Ch[Sc] conformation. Whereas ChQ15R retained the ability to be seeded by Sc or Ca fibres, ChS17R specifically lost the ability to be seeded by Sc (Fig. 2b). Furthermore, under conditions (25 °C) where Ch adopts a mixture of conformations capable of seeding Sc or Ca, ChQ15R and, to a greater extent, ChS17R form seeds that show a strong preference for seeding Ca (Fig. 2c). In principle, this change in specificity could be due to the mutations *per se*, rather than to a change in fibre conformation. We exclude this possibility by showing that fibres formed by seeding Ch with spontaneously polymerized ChQ15R or ChS17R mutants adopted a Ch[Ca] state (Fig. 2d). Moreover, ChQ15R can be forced into a Ch[Sc]-like form when templated by Sc fibres, as evidenced by the seeding specificity and stability of ChQ15R[Sc] (Supplementary Fig. S3).

The hypothesis that differences in prion strains result from different conformations of the prion protein suggests that an *in*

vitro shift in Ch mutant forms should be accompanied by a similar shift in prion strain phenotypes *in vivo*. To test this, we examined the spectrum of prion strains induced by transient overexpression of the various prion domains using yeast (*S. cerevisiae*) in which the genomic SUP35 prion domain was replaced by the Ch sequence (SUP35-Ch yeast). (Yeast strains are denoted by the species of SUP35 prion domain encoded by the genomic SUP35 gene; see Supplementary Table S1.) As previously reported, using an Sc inducer favours the formation of 'weak' Ch prion states (determined by colony colour on low-adenine media as a readout of *ade1* nonsense suppression read-through), whereas Ca overexpression leads to a preponderance (~90%) of strong variants²¹ (Fig. 3a). Induction with Ch resulted in a strain distribution falling between those seen with the Sc or Ca inducers, whereas the ChQ15R and ChS17R mutants show a progressive shift towards strong 'Ch[Ca]-like' strains (Fig. 3a). These data argue that *in vivo* Ch also spontaneously forms both Ch[Sc] and Ch[Ca] conformations, and that the Q15R and S17R mutations specifically disfavour the formation of weak [*PSI*⁺] variants, which we propose are in the Ch[Sc] state.

This shift in strain preference was accompanied by the creation of a transmission barrier *in vivo*. Whereas SUP35-Ch yeast are promiscuous, capable of being converted to the prion state by either Sc or Ca overexpression, SUP35-ChQ15R yeast and, to a greater extent, SUP35-ChS17R yeast specifically lost susceptibility to induction by Sc (Fig. 3b). This transmission barrier was reciprocal, as overexpression of ChQ15R or ChS17R could efficiently induce the prion state in SUP35-Ca yeast, but not in SUP35-Sc yeast (Fig. 3c).

Taking advantage of a set of four glycine-to-alanine mutations that slowed the polymerization of Ca (P.C. and J.S.W., unpublished data), we produced a variant of Ch (ChGA4; Supplementary Fig. S4) with a decreased propensity to form the Ch[Ca] state. *In vivo*, ChGA4 overexpression efficiently induces prion formation in SUP35-Sc yeast, but not in SUP35-Ca yeast — a specificity reciprocal to that seen with ChS17R (Fig. 3d). Similarly, *in vitro*, ChGA4 spontaneously forms amyloid fibres that preferentially seed Sc over Ca. This specificity appears to result from a change in fibre conformation, rather than through a direct effect of the mutations, as fibres generated by seeding Ch with spontaneously polymerized ChGA4 adopted a Ch[Sc] state (Fig. 3e). Taken together, these data argue that the Ch mutations generate transmission barriers both *in vivo* and *in vitro* by specifically disfavoured subsets of Ch amyloid conformations.

We next examined whether point mutations could alter prion conformations/strains adopted by the natural *S. cerevisiae* Sup35 prion domain (Sc), which, like many amyloid-prone proteins³, including another yeast prion protein Ure2p³, is known to form a range of self-propagating states^{2,4}. Whereas transient overexpression of wild-type or the ScQ15R derivative induced a similar spectrum of strong (white or lightly sector) [*PSI*⁺] variants in a SUP35-Sc background (74-D694), induction by the ScS17R mutant strongly favoured the formation of weak strains (mostly red or heavily sector) (Fig. 4a). *In vitro*, we examined the effects of Q15R and S17R mutations on Sc amyloid conformations using a recently described single-fibre assay based on atomic-force microscopy that follows the growth of unlabelled monomers as they add onto the ends of antibody-labelled seeds². Previously, this assay showed that Sc spontaneously formed several distinct fibre types that differed in both polarity and overall rate of growth (Fig. 4b). The range of these fibre types is highly dependent on sequence, as revealed by a marked difference in growth patterns seen between self-seeded Sc, ScQ15R or ScS17R reactions (Fig. 4b). These changes stem from propagating conformational differences within the fibres, as addition of wild-type Sc monomer onto either ScQ15R or ScS17R seeds resulted in single-fibre growth patterns distinct from the pattern of multiple peaks seen when wild-type adds onto wild-type seeds. Interestingly,

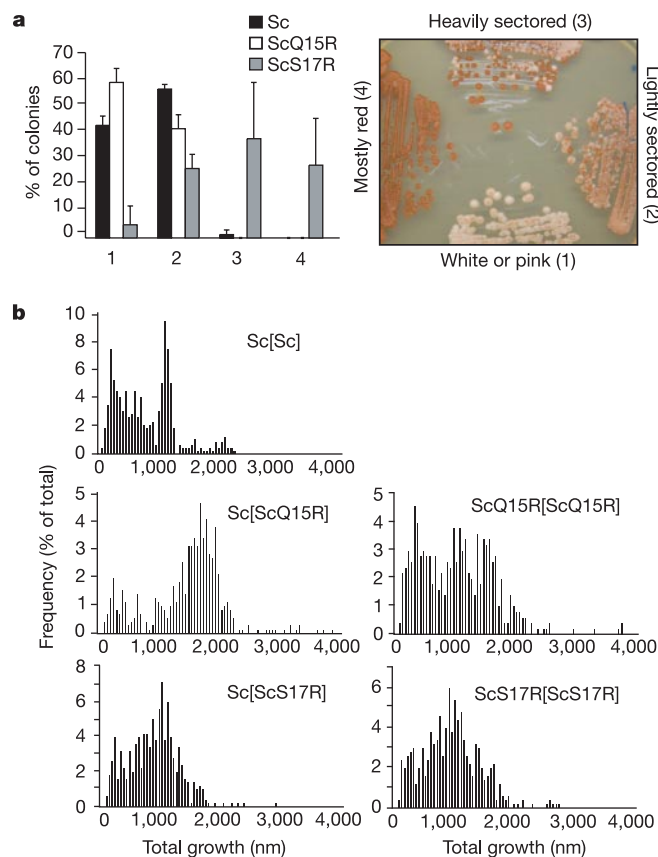


Figure 4 Effects of mutations on the natural *S. cerevisiae* Sup35 prion domain. **a**, Distributions of strain types (left) following induction of [*PSI*⁺] in wild-type yeast (74-D764) by overexpression of Sc (black), ScQ15R (white) or ScS17R (grey) Sup35 prion domains. Examples of each strain category are shown on the right. **b**, Distribution of polymerization rates of the indicated monomers onto individual preformed seeds (indicated in brackets) as determined by an AFM-based assay².

whereas self-seeded ScQ15R reactions yield three discernible fibre types, wild-type reactions seeded by ScQ15R are skewed towards one peak, suggesting that the ScQ15R fibre types vary in their ability to incorporate wild-type protein (Fig. 4b). Thus, both *in vitro* and *in vivo* templating of wild-type protein with the Sc mutants appears to result in propagating changes in the prion state.

The view that a single protein can misfold into multiple distinct self-propagating forms, and that the range of preferred conformations is sensitive to primary structure and perhaps cellular environment^{7,13}, might explain several common features of prion transmission barriers and strains. For example, transmission barriers emerge rapidly relative to the loss of prion-forming ability during evolution^{7,8,10–12}. This could be explained by the fact that each conformation must be self-propagating and that mutations that completely prevent amyloid formation are likely to be rare compared with changes that disfavour only a subset of fibre types or prion strain phenotypes^{23,24} (Figs 3a and 4a). In addition, it has been reported that crossing a transmission barrier can result in a change in the predominant prion strains. Because conformations vary in their specificity, the transmission barrier could act as a sieve, selectively amplifying infectious forms compatible with the recipient prion sequence^{7,25,26}. More generally, the degeneracy in amyloid formation may provide a means of mitigating the toxic effects of protein misfolding: therapeutic strategies designed to promote the formation of non-toxic conformations rather than preventing amyloid formation altogether might be more tractable, and evolutionary pressure might favour sequences that form less toxic conformations when they do misfold. Finally, the failure to create infectious forms of the mammalian prion protein (PrP) *in vitro* self-propagating conformations outside the normal cellular context^{27–29}. □

Methods

Plasmid and yeast-strain construction

Inducer plasmids expressing Sc, Ca and Ch prion domains fused to green fluorescent protein (GFP) under the control of inducible *CUP1* promoters were as described previously^{8,21,22}. Mutant Ch inducers were generated by oligo-directed polymerase-chain-reaction mutagenesis. ScS17R also contained an R98H substitution, and ScQ15R also contained R98H and N93D substitutions. Yeast strains expressing SUP35 with the indicated prion domains were generated by replacing the wild-type chromosomal locus in the parental 74-D694 background as described previously²¹.

In vivo analysis of prion strains and specificity

Efficiency of prion induction was determined by growing yeast strains that harboured appropriate inducer plasmids in synthetic media containing 5 μ M Cu₂SO₄ for 24 h, plating on media lacking adenine, and counting colonies after four days of growth; counts were normalized by plating an appropriate dilution of cells on synthetic complete media and counting after four days growth. Plates shown in Fig. 3d represent eight days of growth, illustrating that specificity of induction is not dependent on duration of growth. Phenotypic strength of induced prion strains was assayed by colony colour²¹ on low-adenine, non-selective YEED media.

Amyloid-formation assays

Prion domains carboxy-terminally tagged with 7 $\times$ (Sc) or 9 $\times$ (all others) poly-histidine were produced as previously described⁸. Conversion to amyloid was monitored by a continuous thioflavin-T-based assay. Concentrated stocks of Sup35 variants stored in denaturant were diluted at least 50-fold into 5 mM KH₂PO₄, 150 mM NaCl, pH 7.4, and 100 μ l were added to 100 μ l of 25 μ M thioflavin-T (Sigma) in 50 mM glycine, pH 8.0. Where indicated, reactions also contained 5% w/w of the appropriate sonicated amyloid seed. Note that the efficiency of the seed is dependent on fibre fragmentation, which can vary from preparation to preparation. However, for a given seed preparation, initial rates of polymerization are linearly proportional to the amount of added seed (S.R.C., unpublished observations), and specificity, as determined by the relative rates of polymerization of two different monomers for a given seed, is therefore independent of the amount or degree of fragmentation of seed. Use of initial rates of polymerization also eliminates the effect of secondary nucleation events such as fibre breakage and *de novo* seed formation. Fluorescence was monitored in a 96-well fluorescence plate reader (Molecular Devices; 442 nm excitation and 485 nm emission). Thioflavin-T assays were carried out at 25 °C and read automatically every minute with 1-s shaking between measurements. Single-filament assays were performed as described³, with the exception that here seeds produced by spontaneous polymerization were used directly, without passage through multiple rounds of polymerization and reseeded. Thermal stability of fibres (Fig. 1c) was determined by incubation of fibres at increasing temperatures (25 °C to 95 °C in 10°

intervals) for 5 min in 1.6% SDS, followed by SDS–polyacrylamide gel electrophoresis (SDS–PAGE) analysis.

Received 15 May; accepted 7 July 2003; doi:10.1038/nature01894.

- Dobson, C. M. The structural basis of protein folding and its links with human disease. *Phil. Trans. R. Soc. Lond. B* **356**, 133–145 (2001).
- DePace, A. H. & Weissman, J. S. Origins and kinetic consequences of diversity in Sup35 yeast prion fibers. *Nature Struct. Biol.* **9**, 389–396 (2002).
- Baxa, U., Speransky, V., Steven, A. C. & Wickner, R. B. Mechanism of inactivation on prion conversion of the *Saccharomyces cerevisiae* Ure2 protein. *Proc. Natl Acad. Sci. USA* **99**, 5253–5260 (2002).
- Glover, J. R. *et al.* Self-seeded fibers formed by Sup35, the protein determinant of [PSI⁺], a heritable prion-like factor of *S. cerevisiae*. *Cell* **89**, 811–819 (1997).
- Kad, N. M., Thomson, N. H., Smith, D. P., Smith, D. A. & Radford, S. E. β_2 -Microglobulin and its deamidated variant, N17D form amyloid fibrils with a range of morphologies *in vitro*. *J. Mol. Biol.* **313**, 559–571 (2001).
- Lashuel, H. A., Hartley, D., Petre, B. M., Walz, T. & Lansbury, P. T. Jr Neurodegenerative disease: amyloid pores from pathogenic mutations. *Nature* **418**, 291 (2002).
- Collinge, J. Prion diseases of humans and animals: their causes and molecular basis. *Annu. Rev. Neurosci.* **24**, 519–550 (2001).
- Santoso, A., Chien, P., Osherovich, L. Z. & Weissman, J. S. Molecular basis of a yeast prion species barrier. *Cell* **100**, 277–288 (2000).
- Come, J. H., Fraser, P. E. & Lansbury, P. T. Jr A kinetic model for amyloid formation in the prion diseases: importance of seeding. *Proc. Natl Acad. Sci. USA* **90**, 5959–5963 (1993).
- Chernoff, Y. O. *et al.* Evolutionary conservation of prion-forming abilities of the yeast Sup35 protein. *Mol. Microbiol.* **35**, 865–876 (2000).
- Kushnirov, V. V., Kochneva-Pervukhova, N. V., Chechenova, M. B., Frolova, N. S. & Ter-Avanesyan, M. D. Prion properties of the Sup35 protein of yeast *Pichia methanolicus*. *EMBO J.* **19**, 324–331 (2000).
- Nakayashiki, T., Ebihara, K., Bannai, H. & Nakamura, Y. Yeast [PSI⁺] “prions” that are cross-transmissible and susceptible beyond a species barrier through a quasi-prion state. *Mol. Cell* **7**, 1121–1130 (2001).
- Uptain, S. M. & Lindquist, S. Prions as protein-based genetic elements. *Annu. Rev. Microbiol.* **56**, 703–741 (2002).
- King, C. Y. *et al.* Prion-inducing domain 2–114 of yeast Sup35 protein transforms *in vitro* into amyloid-like filaments. *Proc. Natl Acad. Sci. USA* **94**, 6618–6622 (1997).
- Derkatch, I. L., Chernoff, Y. O., Kushnirov, V. V., Inge-Vechtomov, S. G. & Liebman, S. W. Genesis and variability of [PSI⁺] prion factors in *Saccharomyces cerevisiae*. *Genetics* **144**, 1375–1386 (1996).
- Kushnirov, V. V., Kryndushkin, D. S., Boguta, M., Smirnov, V. N. & Ter-Avanesyan, M. D. Chaperones that cure yeast artificial [PSI⁺] and their prion-specific effects. *Curr. Biol.* **10**, 1443–1446 (2000).
- Zhou, P. *et al.* The yeast non-Mendelian factor [ETA⁺] is a variant of [PSI⁺], a prion-like form of release factor eRF3. *EMBO J.* **18**, 1182–1191 (1999).
- Kochneva-Pervukhova, N. V. *et al.* [PSI⁺] prion generation in yeast: characterization of the ‘strain’ difference. *Yeast* **18**, 489–497 (2001).
- Uptain, S. M., Sawicki, G. J., Caughey, B. & Lindquist, S. Strains of [PSI⁺] are distinguished by their efficiencies of prion-mediated conformational conversion. *EMBO J.* **20**, 6236–6245 (2001).
- King, C. Y. Supporting the structural basis of prion strains: induction and identification of [PSI⁺] variants. *J. Mol. Biol.* **307**, 1247–1260 (2001).
- Chien, P. & Weissman, J. S. Conformational diversity in a yeast prion dictates its seeding specificity. *Nature* **410**, 223–227 (2001).
- DePace, A. H., Santoso, A., Hillner, P. & Weissman, J. S. A critical role for amino-terminal glutamine/asparagine repeats in the formation and propagation of a yeast prion. *Cell* **93**, 1241–1252 (1998).
- Barron, R. M. *et al.* Changing a single amino acid in the N-terminus of murine PrP alters TSE incubation time across three species barriers. *EMBO J.* **20**, 5070–5078 (2001).
- Mastrianni, J. A. *et al.* Inherited prion disease caused by the V210I mutation: transmission to transgenic mice. *Neurology* **57**, 2198–2205 (2001).
- Peretz, D. *et al.* A change in the conformation of prions accompanies the emergence of a new prion strain. *Neuron* **34**, 921–932 (2002).
- Bartz, J. C., Bessen, R. A., McKenzie, D., Marsh, R. F. & Aiken, J. M. Adaptation and selection of prion protein strain conformations following interspecies transmission of transmissible mink encephalopathy. *J. Virol.* **74**, 5542–5547 (2000).
- Hill, A. E., Antoniou, M. & Collinge, J. Protease-resistant prion protein produced *in vitro* lacks detectable infectivity. *J. Gen. Virol.* **80**, 11–14 (1999).
- Baskakov, I. V., Legname, G., Baldwin, M. A., Prusiner, S. B. & Cohen, F. E. Pathway complexity of prion protein assembly into amyloid. *J. Biol. Chem.* **277**, 21140–21148 (2002).
- Post, K., Brown, D. R., Groschup, M., Kretzschmar, H. A. & Riesner, D. Neurotoxicity but not infectivity of prion proteins can be induced reversibly *in vitro*. *Arch. Virol. Suppl.* **16**, 265–273 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank H. Wille, J. Hood-DeGrenier and members of the Weissman and Lim lab for discussion and critical reading of the manuscript. P.C. and S.R.C. were supported by National Science Foundation Graduate Fellowships and the ARCS (Achievement Rewards for College Scientists) foundation (P.C.). A.H.D. was supported by a Howard Hughes Medical Institute predoctoral fellowship. Funding was also provided by Howard Hughes Medical Institute, The David and Lucile Packard Foundation and the National Institutes of Health.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.S.W. (jswl@itsa.ucsf.edu).

BAD and glucokinase reside in a mitochondrial complex that integrates glycolysis and apoptosis

Nika N. Danial¹, Colette F. Gramm¹, Luca Scorrano^{1*}, Chen-Yu Zhang², Stefan Krauss², Ann M. Ranger¹, Sandeep Robert Datta³, Michael E. Greenberg³, Lawrence J. Licklider⁴, Bradford B. Lowell², Steven P. Gygi⁴ & Stanley J. Korsmeyer¹

¹Howard Hughes Medical Institute, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115, USA

²Division of Endocrinology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts 02215, USA

³Division of Neuroscience, Children's Hospital, Department of Neurobiology, Harvard Medical School, Boston, Massachusetts 02115, USA

⁴Department of Cell Biology, Taplin Biological Mass Spectrometry Facility, Harvard Medical School, Boston, Massachusetts 02115, USA

* Present address: Venetian Institute of Molecular Medicine, Padova 35129, Italy

Glycolysis and apoptosis are considered major but independent pathways that are critical for cell survival^{1–4}. The activity of BAD, a pro-apoptotic BCL-2 family member, is regulated by phosphorylation in response to growth/survival factors^{5–8}. Here we undertook a proteomic analysis to assess whether BAD might also participate in mitochondrial physiology. In liver mitochondria, BAD resides in a functional holoenzyme complex together with protein kinase A⁷ and protein phosphatase 1 (PP1) catalytic units⁹, Wiskott–Aldrich family member WAVE-1 as an A kinase anchoring protein¹⁰, and glucokinase (hexokinase IV)¹¹. BAD is required to assemble the complex in that *Bad*-deficient hepatocytes lack this complex, resulting in diminished mitochondria-based glucokinase activity and blunted mitochondrial respiration in response to glucose. Glucose deprivation results in dephosphorylation of BAD, and BAD-dependent cell death. Moreover, the phosphorylation status of BAD helps regulate glucokinase activity. Mice deficient for BAD or bearing a non-phosphorylatable BAD(3SA) mutant¹² display abnormal glucose homeostasis including profound defects in glucose tolerance. This combination of proteomics, genetics and physiology indicates an unanticipated role for BAD in integrating pathways of glucose metabolism and apoptosis.

Multidomain pro-apoptotic BCL-2 members BAX and BAK constitute a requisite gateway to the mitochondrial pathway of apoptosis, and their allosteric activation results in permeabilization of the outer mitochondrial membrane^{4,13}. BH3-only molecules including BAD, BID, NOXA and BIM are connected to proximal death and survival signals, and can result in the activation of BAX and BAK or can be sequestered by anti-apoptotic BCL-2 and BCL-X_L¹⁴. Kinases that phosphorylate and inactivate BAD have been shown to localize to mitochondrial membranes, and depending on cellular context include mitochondria-tethered protein kinase A (PKA) for S112 and S155 sites in response to interleukin-3 (IL-3), and a mitochondria-based component of p70S6 kinase for S136 after insulin-like growth factor (IGF)-I signalling (refs 6–8).

To assess whether BAD and perhaps enzymes responsible for its modification reside in higher-order protein complexes at the mitochondria, we performed gel filtration chromatography. Conditions for fractionation were established using mitochondria purified from the haematopoietic cell line FL5.12 expressing BAD and BCL-X_L (Supplementary Fig. 1). To examine the fractionation profile of endogenous BAD within normal tissue, Percoll centrifugation was used to separate mouse liver mitochondria from other cellular membranes. Purified mitochondria were solubilized in

CHAPS, a zwitterionic detergent that does not alter the conformation of BCL-2 family proteins¹⁵. We initially used mild conditions (1 mM CHAPS) to keep peripheral protein complexes intact, followed by Superose 6 gel filtration (Fig. 1a). Immunoblots of gel-excluded fractions revealed that BAD_L (28 kDa) and BAD_S (23 kDa) isoforms were present within complexes of 25–669 kDa in mouse liver mitochondria and that the BAD_S species predominated (Fig. 1b). The fractionation properties of BAD remained the same when mitochondria were solubilized in 1, 6 or 15 mM (1%) CHAPS (critical micelle concentration about 6–10 mM) (Supplementary Fig. 2). In 1–15 mM CHAPS the fractionation properties of VDAC, the outer mitochondrial membrane import receptor Tom20, the import channel Tom40 and the mitochondrial chaperone Hsp60 are distinct and consistent with published observations^{16,17}.

BAD-containing fractions from gel filtration were separated under non-denaturing conditions, revealing multiple silver-stained protein complexes (Fig. 1c). Immunoblots of these second dimension native gels revealed BAD immunoreactivity within distinct complexes in fractions C to E2 from liver mitochondria (Fig. 1d). Gel-excluded proteins (first dimension), which were fractionated into native complexes under non-denaturing conditions (second dimension), were subsequently separated into their individual components (third dimension) by placing a native gel slice horizontal as a stack above an SDS–polyacrylamide gel electrophoresis (PAGE) denaturing gel (Fig. 1a). Silver stains of third dimension gels (Fig. 1e) revealed the individual protein constituents of each native complex (Fig. 1c). Third dimension gels were also transferred and immunoblotted, confirming the presence of 23-kDa BAD_S, but not BAD_L, in an approximately 232-kDa complex (Fig. 2a). Five potential component proteins of a roughly 232-kDa BAD-containing complex were consistently present in silver-stained third dimension gels of fractions D to E2 (Fig. 1e, red arrows). Moreover, immunoblots using an antibody specific for the phosphorylated S112 site of BAD indicate that at least a portion of BAD_S present within this complex is phosphorylated (data not shown). The absence of an immunoreactive complex in *Bad*-deficient liver mitochondria (Fig. 2a) confirmed the specificity of the BAD-containing complex. In contrast, manganese SOD localized in complex(es) ranging from about 158 to 232 kDa in *Bad*^{−/−} as well as *Bad*^{+/+} mitochondria, indicating that only BAD-specific complexes were disrupted in the *Bad* knockout liver (Fig. 2a).

The approximately 232-kDa native protein complexes isolated from second dimension gels (Fig. 1c) and individual protein species excised from third dimension gels (Fig. 1e) were sequenced by liquid chromatography tandem mass spectrometry (LC-MS/MS), which identified a peptide that is 94% identical to the catalytic subunit of human PP1α and rat PP1α serine/threonine phosphatase. Its precise identity as the 36-kDa catalytic subunit (PP1_C) was confirmed by immunoblot of a third dimension gel (Fig. 2a). The correctly sized BAD_S, PP1_C and PKA_C vertically align on third dimension gels, supporting their presence together within a complex of roughly 232 kDa (Fig. 2a).

A second approach to assess whether candidate proteins resided together in a complex used immunoprecipitation of fractions from FL5.12 mitochondria, revealing that BAD_S and PKA_C precipitated together with PP1_C (Supplementary Fig. 1c). As a third assessment of whether candidate proteins associate, we isolated PP1_C-interacting proteins by microcystin affinity chromatography, which efficiently purifies PP1_C owing to its high-capacity binding to this phosphatase⁹. Microcystin-Sepharose affinity pull-downs revealed multiple proteins, and immunoblots confirmed the purification of PP1_C (Fig. 2b). Another closely related phosphatase with microcystin-binding capacity, namely PP2A, was not present (Supplementary Fig. 3). PKA_C and BAD_S also co-purified with the captured PP1 from wild-type mitochondria (Fig. 2b), supporting all other evidence showing their presence together in a mitochondrial complex.

Confocal microscopy of mouse embryonic fibroblasts (MEFs) demonstrated that PP1_C localizes together with Tom20 (Fig. 2d). Submitochondrial fractionation of mouse liver mitochondria confirmed that PP1 is detected in the mitochondrial outer membrane fraction (data not shown). The presence of PP1 at mitochondria was not expected¹⁸. Our data indicate that a portion of phosphorylated BAD is present at mitochondria, supporting opposing kinase and phosphatase activities as part of the same BAD-containing multi-enzyme complex.

To test the role of PP1 in BAD dephosphorylation we used a PP1-blocking peptide known to inhibit subcellular targeting of PP1 and its dephosphorylation activity on substrates¹⁹. Addition of the R/K-X-V/I-X-F motif containing blocking peptide to mitochondria of FL5.12 cells, prepared immediately after their withdrawal from IL-3, prevented the dephosphorylation of BAD, but the mutant F68A peptide did not (Fig. 2e; see also Supplementary Fig. 4). Of note, the requirement for $>10^{-6}$ M okadaic acid to inhibit the phosphatase activity for mitochondrial BAD indicates the partici-

pation of PP1, not PP2A (Supplementary Fig. 4). These approaches support a model in which PP1 is specifically targeted to mitochondria where it resides in an enzyme complex responsible for the dynamic dephosphorylation of BAD.

The presence of PKA and PP1 suggested that the complex might possess an A kinase anchoring protein (AKAP). Whereas D-AKAP1 had been noted at the mitochondria of FL5.12 cells⁷, LC-MS/MS of the approximately 232-kDa native complexes from liver identified WAVE-1, a member of Wiskott-Aldrich family proteins²⁰. In addition to regulating actin polymerization, WAVE-1 functions as an AKAP, binding the PKA regulatory subunit (RII)¹⁰. An RII overlay as well as immunoblots performed on a third dimension gel recognized approximately 232, 134 and 67-kDa species, implicating WAVE-1 as the AKAP within these mitochondrial complexes (Fig. 2f). Moreover, WAVE-1 aligned vertically with BAD_S and PP1_C on a third dimension gel, evidence indicating that they resided in a BAD-containing complex (Fig. 2f). Further evidence that WAVE-1 resided within this holoenzyme complex was provided by its presence in microcystin pull-downs of PP1_C (Fig. 2b; see also Supplementary Fig. 3). Immunoprecipitation of mitochondrial proteins solubilized in 15 mM CHAPS revealed that WAVE-1 precipitated together with BAD, providing a third approach to confirm their association (Fig. 2c). Immunoelectron microscopy using an antibody to WAVE-1 revealed prominent decoration of the outer mitochondrial membrane (Fig. 2g). Immunolocalization studies using confocal microscopy indicated a portion of WAVE-1 localized with Tom20 at the mitochondria (Supplementary Fig. 5).

To assess whether the BAD-containing complex at mitochondria actually required BAD for its assemblage, we performed a similar first, second and third dimension analysis on the mitochondria from livers of *Bad*-deficient mice. Immunoblots of third dimension gels from fractions D through to E2 failed to detect any PP1_C, PKA_C, WAVE-1 or BAD (Fig. 2a, f). An examination of multiple fractions indicates that deletion of BAD results in disassembly of the entire protein complex rather than a shift to a slightly smaller-sized complex. This suggests that BAD also functions as a scaffold to nucleate the proper architecture of this complex, as well as being a substrate for the kinase-phosphatase pair localized to this microenvironment.

The components BAD_S (23 kDa), PKA_C (41 kDa), PP1_C (36 kDa) and WAVE-1 (85-kDa isoform) do not account for the approximately 50-kDa protein noted by silver stain on the third dimension gel of the approximately 232-kDa core complex (Fig. 1e). LC-MS/MS analysis of the PP1-associated proteins captured from liver mitochondrial lysates identified glucokinase (Fig. 2b; see also Supplementary Fig. 3), a 52-kDa protein whose expression pattern is mainly restricted to liver and β -islet cells^{11,21}. An immunoblot of the microcystin pull-down confirmed that glucokinase precipitated together with PP1_C, BAD_S, PKA_C and WAVE-1 (Fig. 2b, see also Supplementary Fig. 3). Moreover, an immunoprecipitation of glucokinase from 15 mM CHAPS-solubilized mitochondria also precipitated together with WAVE-1, BAD_S and PKA_C (Fig. 2c). Further evidence that glucokinase was a component of the approximately 232-kDa complex was provided by immunoblots of third dimension gels where glucokinase aligned vertically with BAD_S, PKA_C and PP1_C (Fig. 2a).

Absence of the approximately 232-kDa complex containing glucokinase in *Bad*^{-/-} cells prompted an assessment of the physiological consequence of losing the mitochondria-tethered component of this enzyme. The rate of organelle-based glucokinase activity was severely blunted in *Bad*^{-/-} compared with wild-type liver mitochondria (Fig. 3a, $P < 0.01$), indicating that a localized enzymatic defect resulted from the loss of a focused glucokinase complex. An immunoblot comparing purified mitochondria and cytosol from hepatocytes confirmed a decrease in glucokinase protein in *Bad*^{-/-} mitochondria (Fig. 3b).

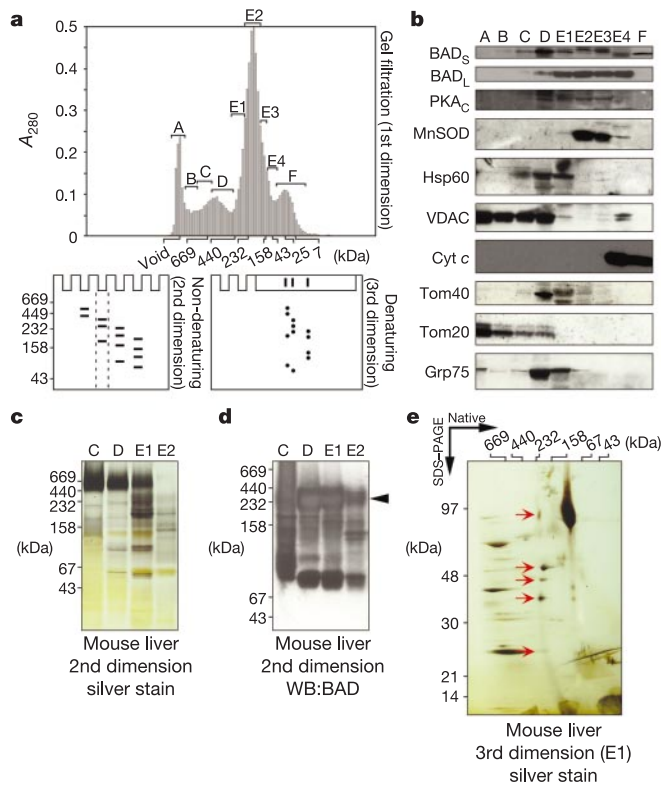


Figure 1 Characterization of mitochondrial complexes containing BAD. **a**, Schematics depicting the first, second and third dimension separation of mitochondrial protein complexes. Top panel, elution profile of mitochondrial proteins. Mitochondria isolated from mouse liver were solubilized in 1 mM CHAPS, subjected to gel filtration on Superose 6, and 200- μ l fractions were collected. Protein concentration is plotted to identify elution peaks. A_{280} , absorbance at 280 nm. **b**, Fifty micrograms of pooled fractions were loaded onto an 8–16% gradient SDS-PAGE and immunoblotted with the indicated antibodies. MnSOD, manganese SOD. **c**, **e**, Representative silver stain of second (**c**) and third (**e**) dimension gel fractionation of mitochondrial complexes, that were subsequently subjected to LC-MS/MS. Arrows in **e** indicate five components of an approximately 232-kDa complex consistently noted on third dimension analysis. **d**, Native BAD-containing complexes of mouse liver mitochondria. One hundred and fifty micrograms of gel-filtrated liver mitochondrial fractions were subjected to non-denaturing electrophoresis (second dimension). Proteins were transferred and immunoblotted (WB) with an anti-BAD antibody. Arrow indicates the most prominent BAD complex.

Although glucokinase activity at the surface of *Bad*^{-/-} mitochondria was markedly diminished (Fig. 3a), most of the glucokinase resides in the cytosol (Fig. 3b). This prompted us to ask whether intact cells from these mice would exhibit a specific defect in the use of glucose as a respiratory substrate. Addition of glucose to wild-type hepatocytes increased the rate of oxygen consumption, whereas respiration driven by glucose in *Bad*^{-/-} hepatocytes was significantly compromised, as depicted by the slope of the oxygen consumption curve (Fig. 3c). Consistent with this, the percentage increase in ATP after treatment with glucose was also decreased in *Bad*^{-/-} hepatocytes (data not shown). If the purpose of BAD-dependent, mitochondrial-tethered glucokinase is to provide this organelle directly with metabolic intermediates to drive mitochondrial processes including respiration, then the respiration of *Bad*^{-/-} hepatocytes might be corrected by substituting pyruvate, a downstream product of glycolysis. Addition of exogenous pyruvate to digitonin-permeabilized *Bad*^{-/-} hepatocytes restored the oxygen consumption rate to that of wild-type cells (Fig. 3c, d). Moreover, respiration in response to an alternative carbon source, fructose, was also equivalent in *Bad*^{-/-} cells (Fig. 3d). Thus, the respiratory complexes in *Bad*^{-/-} liver seem intact, strongly supporting the role of a BAD-dependent multiprotein complex in the usage of glucose at mitochondria.

As glucose metabolism is linked to cell survival^{2,3}, we assessed BAD phosphorylation in the presence or absence of glucose. BAD S112 was phosphorylated when glucose was present but not when it was absent from the medium of a hepatocyte cell line (Fig. 3e).

Addition of glucose to purified liver mitochondria induced the phosphorylation of BAD (Fig. 3e). Moreover, the absence of BAD markedly protected liver cells from apoptosis when deprived of glucose (Fig. 3f).

Mutations in glucokinase have been associated with maturity-onset type 2 diabetes of the young¹¹. Similar defects in glucose regulation have been recapitulated in mice with genetic alterations in the glucokinase gene^{21,22}, prompting us to examine whether BAD has an effect on glucose homeostasis. Blood glucose measurements in fed and fasted mice revealed a substantial fasting hyperglycaemia (28% increase, *P* < 0.05, Fig. 3g) as well as a fed hyperinsulinaemia (not shown) in *Bad*^{-/-} mice. After intraperitoneal injection of glucose (1 g glucose per kg mouse), *Bad*^{-/-} mice showed a significant defect in the rate of blood glucose clearance (Fig. 3h).

As glucose could regulate the phosphorylation of BAD in hepatocytes (Fig. 3e), we tested whether the phosphorylation status of BAD affected the mitochondrial complex. We examined hepatocytes from a 'knockin' mouse bearing a non-phosphorylatable BAD(3SA) molecule in which the three regulatory serines (S112, S136 and S155) were converted to alanine¹². Although BAD protein is required to nucleate this mitochondrial complex (Fig. 2a), the phosphorylation of BAD is not required as BAD(3SA) liver contained the multiprotein complex, as demonstrated by third dimension analysis (Fig. 4a) and microcystin pull-down (Fig. 2b). However, despite normal amounts of mitochondrial glucokinase protein, the glucokinase activity was markedly blunted in

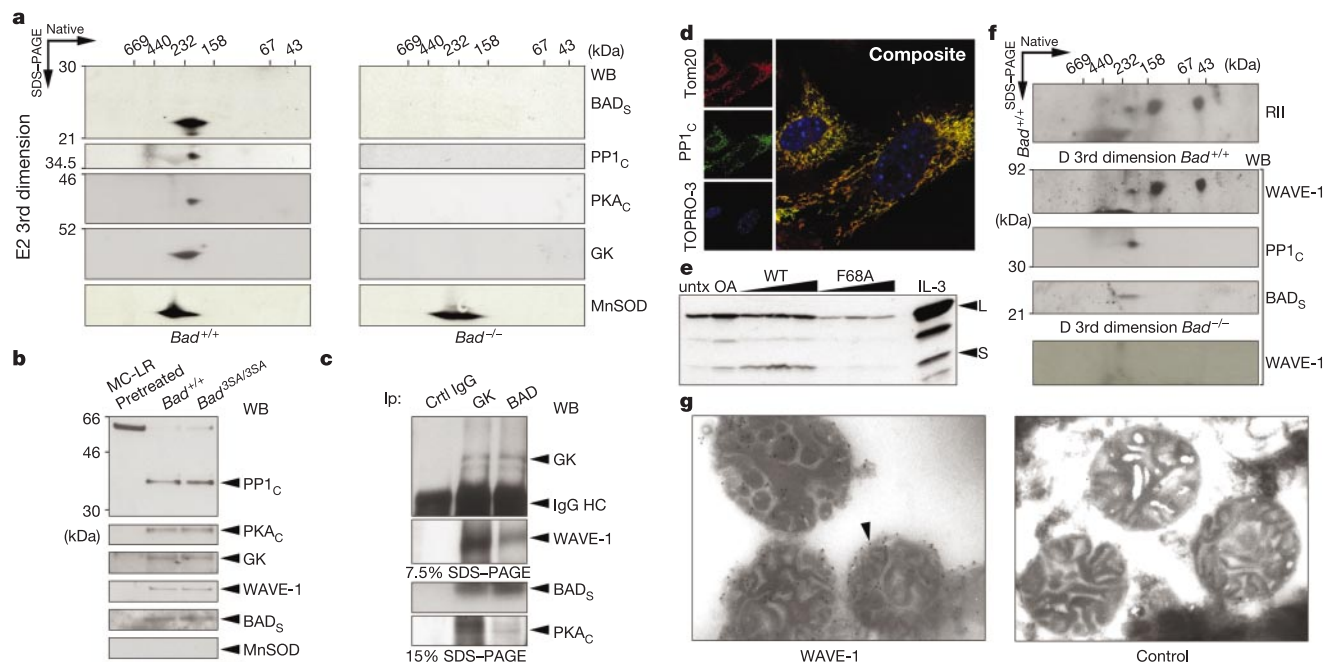


Figure 2 Components of a mitochondrial BAD complex. **a**, Immunoblot of third dimension gel of a mitochondrial protein complex from the liver of *Bad*^{+/+} and *Bad*^{-/-} mice. GK, glucokinase. **b**, Microcystin affinity purification of PP1-interacting proteins in mouse liver mitochondria. Mitochondria solubilized in 15 mM CHAPS were incubated with microcystin-coupled Sepharose (MC-Sepharose) beads and bound proteins were resolved by SDS-PAGE and immunoblotted (WB) with the indicated antibodies. Pretreatment of mitochondrial lysates with microcystin serves as a control. **c**, Co-immunoprecipitation of BAD_s, glucokinase, WAVE-1 and PKA_C. Mouse liver mitochondrial proteins solubilized in 15 mM CHAPS were immunoprecipitated (IP) with an antibody to BAD or glucokinase, immunoprecipitates were resolved on 7.5% or 15% SDS-PAGE gels and immunoblots were developed with the indicated antibodies. IgG HC, immunoglobulin-γ heavy chain.

d, Immunolocalization of PP1 to mitochondria by confocal microscopy of MEFs revealing co-localization with Tom20. **e**, PP1 targeting inhibitor prevents BAD dephosphorylation. Purified mitochondria from FL5.12 cells deprived of IL-3 were incubated with 10–1,000 nM wild-type (WT) or mutant (F68A) peptide¹⁹, and BAD phosphorylation on S112 was examined. OA, okadaic acid. **f**, RII-binding activity within the BAD complex. A third dimension gel of filtrate fraction D from wild-type liver mitochondria was analysed by an RII overlay assay¹⁹ or immunoblot. An immunoblot of the third dimension gel of fraction D from *Bad*^{-/-} liver mitochondria was also developed for WAVE-1. **g**, Immunogold decoration of mitochondria by a WAVE-1 antibody (arrowhead) as detected by immunoelectron microscopy of thin sections from purified liver mitochondria.

BAD(3SA) mitochondria (Fig. 4b, $P < 0.001$). Moreover, mitochondrial glucokinase activity was substantially inhibited by a PKA inhibitor (H-89) (Fig. 4c) or a PKA-specific inhibitory peptide (PKI) (not shown), both of which blocked S112 phosphorylation⁷. Finally, we examined *Bad*^{3SA/3SA} mice *in vivo*, revealing a com-

parable defect in glucose homeostasis to that in *Bad*^{-/-} mice. This includes significant fasting hyperglycaemia (Fig. 4d) as well as a marked defect in blood glucose clearance (Fig. 4e). These findings provide genetic evidence that BAD phosphorylation helps to regulate glucose homeostasis.

Cells depend on the availability of growth/survival factors that characteristically regulate both cellular metabolism and cell survival. Insulin, IGF-I and multiple cytokines transduce signals via phosphatidylinositol-3-OH kinase through AKT and related kinases, which maintain glucose transport (Glut) and recruit hexokinase to mitochondria, stimulating glycolysis and ATP production^{3,23}. Conversely, withdrawal of growth/survival factors decreases the glycolytic rate and O₂ consumption, alters mitochondrial membrane potential, and results in VDAC closure and diminished ADP/ATP exchange^{2,3}. Overexpression of anti-apoptotic BCL-2 and BCL-X_L can block the death of cells deprived of growth/survival factors, but does not directly counter the decline in glycolysis^{24,25}. These observations suggested that BCL-2 family proteins and glucose metabolism influence cell survival through independent mechanisms. However, our studies of *Bad*^{-/-} and non-phosphorylatable BAD(3SA) cells reveal a unique role

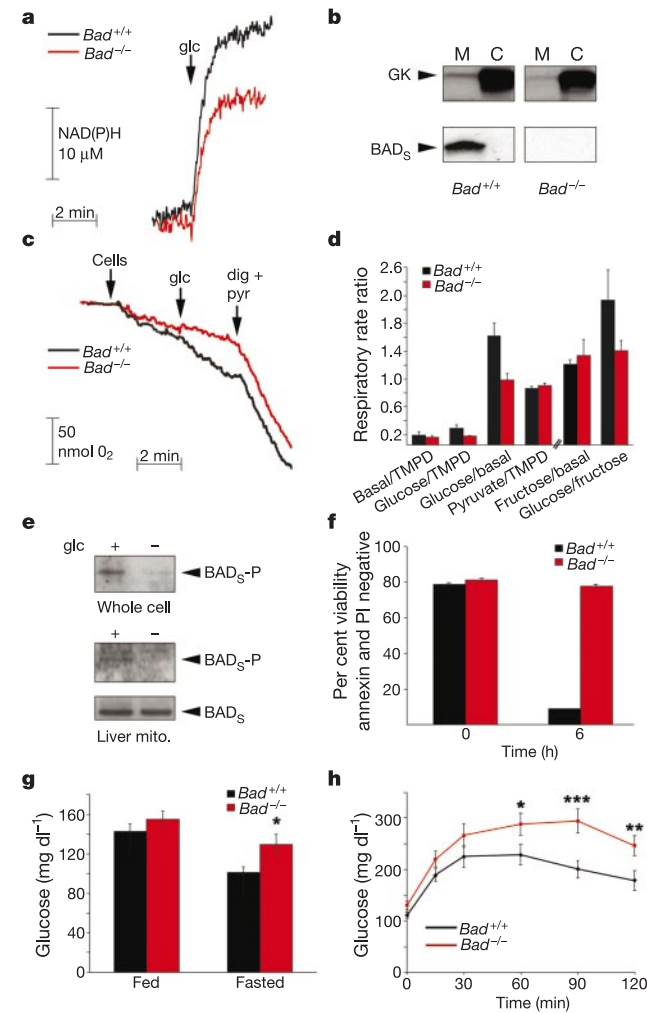


Figure 3 Aberrations in glucokinase activity, glucose homeostasis and glucose-withdrawal-induced cell death in *Bad*^{-/-} mice. **a**, Glucokinase activity in purified liver mitochondria is plotted as glucose-6-phosphate dehydrogenase-driven increase in NAD(P)H fluorescence. Representative trace of multiple independent experiments is shown. **b**, Immunoblot of glucokinase levels in *Bad*^{+/+} and *Bad*^{-/-} mitochondria (M) and cytosol (C) subfractions of liver. **c**, Representative respiratory traces of whole-cell O₂ consumption in response to glucose (glc) or pyruvate-driven respiration comparing *Bad*^{+/+} with *Bad*^{-/-} hepatocytes. Recordings were conducted using a Clark-type oxygen electrode chamber. **d**, Respiratory rate ratios of experiments in **c**. There is a significant difference of *Bad*^{+/+} compared with *Bad*^{-/-} hepatocytes; $P = 0.002$. **e**, Glucose-regulated BAD phosphorylation on S112 (BAD_s-P) in whole-cell lysates of hepatocyte line H4IIE incubated in the presence or absence of glucose (top panel) or in isolated wild-type liver mitochondria treated with glucose *in vitro* (bottom panels). **f**, Viability of cultured hepatocytes shifted from high (22 mM) to low (0.05 mM) glucose medium. **g**, Fed and overnight fasted serum levels of glucose measured in male *Bad*^{+/+} ($n = 5$) and *Bad*^{-/-} ($n = 8$) mice. **h**, Intraperitoneal glucose tolerance test³⁰ plots the averaged values of three independent experiments performed on male *Bad*^{+/+} ($n = 5$) and *Bad*^{-/-} ($n = 8$) mice. Asterisks in **g**, **h** indicate significant differences of *Bad*^{-/-} compared with *Bad*^{+/+} mice. Asterisk, $P < 0.05$; double asterisk, $P < 0.01$; triple asterisk, $P < 0.001$, unpaired two-tailed *t*-test.

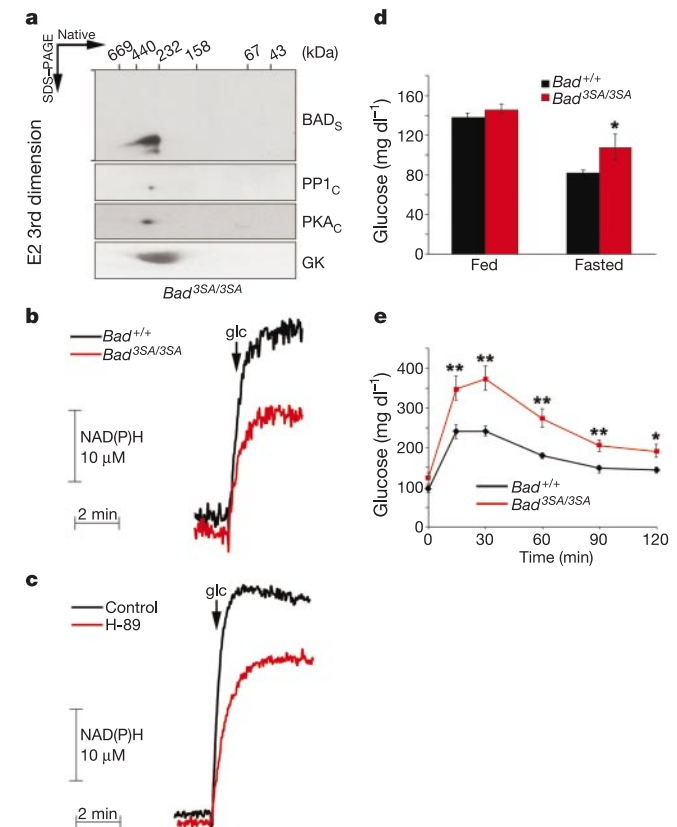


Figure 4 Requirement of BAD phosphorylation in the regulation of mitochondrial glucokinase activity and glucose homeostasis. **a**, The assembly of the mitochondrial BAD complex does not require BAD phosphorylation sites. Third dimension analysis of the BAD complex in *Bad*^{3SA/3SA} liver mitochondria. **b**, Representative plot of mitochondrial glucokinase activity in wild-type compared with *Bad*^{3SA/3SA} liver mitochondria. **c**, Inhibition of mitochondria glucokinase activity by PKA inhibitor H-89. **d**, **e**, Glucose homeostasis in *Bad*^{3SA/3SA} mice. **d**, Fed and overnight fasted serum levels of glucose. **e**, Rate of glucose clearance after intraperitoneal glucose tolerance test measured in male *Bad*^{+/+} ($n = 7$) and *Bad*^{3SA/3SA} ($n = 7$) mice. Asterisks indicate significant differences of *Bad*^{+/+} compared with *Bad*^{3SA/3SA} mice. Asterisk, $P < 0.05$; double asterisk, $P < 0.01$.

for BAD in glucose homeostasis.

The hepatocyte defect and abnormalities in glucose metabolism of *Bad* knockout mice are very similar to a liver-restricted conditional knockout of glucokinase, which demonstrates fasting hyperglycaemia and fed hyperinsulinaemia²¹. This endorses the profound effect of the mitochondria component of glucokinase as essential for proper glucose metabolism. The Michaelis constant (K_m) of glucokinase is close to physiological glucose concentrations, making it an attractive candidate for sensing glucose in this complex. The similarity of *Bad* knockout and knockin mice to diabetic models warrants future, detailed, whole-animal physiological studies, including hyperglycaemic clamps, to define the complete spectrum of systemic changes and compensatory events.

BAD, independent of its phosphorylation status, nucleates a glucokinase-containing complex at the mitochondrion. The *Bad*^{35A} allele provides evidence that phosphorylation of BAD is necessary for maximal glucokinase activity, and endorses a kinase-PTPase pair working together within this complex. Notably, the presence of WAVE-1 that controls actin remodelling may relate to the reported actin-dependent organization of glycolytic enzymes, including glucokinase²⁶, suggesting a rationale for the presence of this particular AKAP in the complex. The striking requirement for BAD in glucose-deprivation-induced death suggests that BAD also serves as an apoptotic sentinel, responding to abnormalities in glucose metabolism. Thus, BH3-only molecules can be integral participants in the pathways that they monitor, ensuring that glycolysis and apoptosis are coordinated. □

Methods

Third dimension analysis of mitochondrial complexes

Mouse liver mitochondria were isolated by standard differential centrifugation (see Supplementary Methods). The procedure described by ref. 27 was adapted as follows. Gel-filtrated mitochondrial proteins were resolved onto 4–12% pre-cast gels (BioRad). Gel lanes were excised and soaked in 2% (w/v) SDS, 0.2 M dithiothreitol (DTT) for 20 min. Each gel strip was layered on top of an 8–16% gradient resolving gel, and a 7% stacking gel poured over and around the native gel slice.

Mass spectrometry

Non-denaturing gels (second dimension; Fig. 1c) or SDS-PAGE gels of microcystin-Sepharose capture assays (Supplementary Fig. 3) were stained either with Coomassie blue or silver stain, and divided horizontally into about ten slices. Resulting gel slices or protein spots from third dimension analysis (Fig. 1e) were individually processed to yield in-gel trypsin digestion products. The resulting peptide mixtures were separated and analysed in an automated system by nanoscale LC-MS/MS on an LCQ-DECA ion trap mass spectrometer (Thermo Finnigan) using the vented-column approach²⁸ (see Supplementary Methods).

Mitochondrial glucokinase assay

Glucokinase activity was measured spectrophotometrically in a glucose-6-phosphate dehydrogenase-driven NAD(P)H fluorescence assay at 340 nm (ref. 29) with the following modifications. Mouse liver mitochondria (1 mg) were resuspended in 500 μ l reaction buffer containing 4 μ g ml⁻¹ glucose-6-phosphate dehydrogenase, 20 mM MgCl₂, 1 mM DTT, 100 mM KCl, 50 mM triethanolamine hydrochloride and 0.1% BSA (see Supplementary Methods).

Microscopy

For confocal microscopy, MEFs were fixed for 20 min at room temperature in PBS containing 3% paraformaldehyde and permeabilized with 0.1% Triton-X100. Incubation with primary and labelled secondary antibodies was performed (see Supplementary Methods for details).

Hepatocyte viability assays

Isolated primary hepatocytes (see Supplementary Methods for details) were resuspended at 6×10^4 cells ml⁻¹ in DMEM high glucose medium (22 mM) supplemented with 10 ng ml⁻¹ HGF (R&D Systems) or shifted to medium containing HGF but low (0.05 mM) glucose. Viability was assessed by flow cytometric analysis of Annexin-V and propidium iodide staining (Bio Vision).

Mitochondrial respiration assay

Isolated hepatocytes were resuspended in PBS at 5×10^6 cells ml⁻¹, transferred to a Clark-type oxygen electrode chamber (Hansatech), and oxygen consumption recordings were started (see Supplementary Methods for details).

Received 20 March; accepted 27 May 2003; doi:10.1038/nature01825.

1. Raff, M. C. Social controls on cell survival and cell death. *Nature* **356**, 397–400 (1992).

- Vander Heiden, M. G. *et al.* Growth factors can influence cell growth and survival through effects on glucose metabolism. *Mol. Cell. Biol.* **21**, 5899–5912 (2001).
- Gottlob, K. *et al.* Inhibition of early apoptotic events by Akt/PKB is dependent on the first committed step of glycolysis and mitochondrial hexokinase. *Genes Dev.* **15**, 1406–1418 (2001).
- Wang, X. The expanding role of mitochondria in apoptosis. *Genes Dev.* **15**, 2922–2933 (2001).
- Zha, J., Harada, H., Yang, E., Jockel, J. & Korsmeyer, S. J. Serine phosphorylation of death agonist BAD in response to survival factor results in binding to 14-3-3 not BCL-X(L). *Cell* **87**, 619–628 (1996).
- Datta, S. R. *et al.* 14-3-3 proteins and survival kinases cooperate to inactivate BAD by BH3 domain phosphorylation. *Mol. Cell* **6**, 41–51 (2000).
- Harada, H. *et al.* Phosphorylation and inactivation of BAD by mitochondria-anchored protein kinase A. *Mol. Cell* **3**, 413–422 (1999).
- Harada, H., Andersen, J. S., Mann, M., Terada, N. & Korsmeyer, S. J. p70S6 kinase signals cell survival as well as growth, inactivating the pro-apoptotic molecule BAD. *Proc. Natl Acad. Sci. USA* **98**, 9666–9670 (2001).
- Moorhead, G., MacKintosh, C., Morrice, N. & Cohen, P. Purification of the hepatic glycogen-associated form of protein phosphatase-1 by microcystin-Sepharose affinity chromatography. *FEBS Lett.* **362**, 101–105 (1995).
- Westphal, R. S., Soderling, S. H., Alto, N. M., Langeberg, L. K. & Scott, J. D. Scar/WAVE-1, a Wiskott-Aldrich syndrome protein, assembles an actin-associated multi-kinase scaffold. *EMBO J.* **19**, 4589–4600 (2000).
- Postic, C., Shiota, M. & Magnuson, M. A. Cell-specific roles of glucokinase in glucose homeostasis. *Recent Prog. Horm. Res.* **56**, 195–217 (2001).
- Datta, S. R. *et al.* Survival factor-mediated BAD phosphorylation raises the mitochondrial threshold for apoptosis. *Dev. Cell* **3**, 631–643 (2002).
- Wei, M. C. *et al.* Proapoptotic BAX and BAK: a requisite gateway to mitochondrial dysfunction and death. *Science* **292**, 727–730 (2001).
- Cheng, E. H. *et al.* BCL-2, BCL-X(L) sequester BH3 domain-only molecules preventing BAX- and BAK-mediated mitochondrial apoptosis. *Mol. Cell* **8**, 705–711 (2001).
- Hsu, Y. T. & Youle, R. J. Bax in murine thymus is a soluble monomeric protein that displays differential detergent-induced conformations. *J. Biol. Chem.* **273**, 10777–10783 (1998).
- Krimmer, T. *et al.* Biogenesis of porin of the outer mitochondrial membrane involves an import pathway via receptors and the general import pore of the TOM complex. *J. Cell Biol.* **152**, 289–300 (2001).
- Abdul, K. M. *et al.* Functional analysis of human metaxin in mitochondrial protein import in cultured cells and its relationship with the Tom complex. *Biochem. Biophys. Res. Commun.* **276**, 1028–1034 (2000).
- Ayllon, V., Martinez, A. C., Garcia, A., Cayla, X. & Rebollo, A. Protein phosphatase 1 α is a Ras-activated Bad phosphatase that regulates interleukin-2 deprivation-induced apoptosis. *EMBO J.* **19**, 2237–2246 (2000).
- Westphal, R. S. *et al.* Regulation of NMDA receptors by an associated phosphatase-kinase signaling complex. *Science* **285**, 93–96 (1999).
- Machesky, L. M. & Insall, R. H. Scar1 and the related Wiskott-Aldrich syndrome protein, WASP, regulate the actin cytoskeleton through the Arp2/3 complex. *Curr. Biol.* **8**, 1347–1356 (1998).
- Postic, C. *et al.* Dual roles for glucokinase in glucose homeostasis as determined by liver and pancreatic beta cell-specific gene knock-outs using Cre recombinase. *J. Biol. Chem.* **274**, 305–315 (1999).
- Bali, D. *et al.* Animal model for maturity-onset diabetes of the young generated by disruption of the mouse glucokinase gene. *J. Biol. Chem.* **270**, 21464–21467 (1995).
- Cho, H. C. *et al.* Insulin resistance and a diabetes mellitus-like syndrome in mice lacking the protein kinase Akt2 (PKB β). *Science* **292**, 1728–1731 (2001).
- Plas, D. R., Talapatra, S., Edinger, A. L., Rathmell, J. C. & Thompson, C. B. Akt and Bcl-XL promote growth factor-independent survival through distinct effects on mitochondrial physiology. *J. Biol. Chem.* **276**, 12041–12048 (2001).
- Garland, J. M. & Halestrap, A. Energy metabolism during apoptosis. Bcl-2 promotes survival in hematopoietic cells induced to apoptose by growth factor withdrawal by stabilizing a form of metabolic arrest. *J. Biol. Chem.* **272**, 4680–4688 (1997).
- Murata, T. *et al.* Co-localization of glucokinase with actin filaments. *FEBS Lett.* **406**, 109–113 (1997).
- Dekker, P. J. *et al.* The Tim core complex defines the number of mitochondrial translocation contact sites and can hold arrested preproteins in the absence of matrix Hsp70-Tim44. *EMBO J.* **16**, 5408–5419 (1997).
- Licklider, L. J., Thoreen, C. C., Peng, J. & Gygi, S. P. Automation of nanoscale microcapillary liquid chromatography-tandem mass spectrometry with a vented column. *Anal. Chem.* **74**, 3076–3083 (2002).
- Niswender, K. D. *et al.* Cell-specific expression and regulation of a glucokinase gene locus transgene. *J. Biol. Chem.* **272**, 22564–22569 (1997).
- Zhang, C. Y. *et al.* Uncoupling protein-2 negatively regulates insulin secretion and is a major link between obesity, beta cell dysfunction, and type 2 diabetes. *Cell* **105**, 745–755 (2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank M. Ryan, G. Shore, J. Scott, M. Magnuson and B. Spiegelman for reagents; M. Ryan and J. Opferman for technical advice; B. Kahn and O. Peroni for discussion; S. Wade, J. Fisher and J. Sturgill for animal care; U. Madueke for technical assistance; and E. Smith for manuscript preparation. N.N.D. is a recipient of the Cancer Research Fund of Damon Runyon Foundation fellowship. This work is supported in part by a NIH grant.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.J.K. (stanley_korsmeyer@dfci.harvard.edu).

Artemisinins target the SERCA of *Plasmodium falciparum*

U. Eckstein-Ludwig¹, R. J. Webb¹, I. D. A. van Goethem², J. M. East², A. G. Lee², M. Kimura³, P. M. O'Neill⁴, P. G. Bray⁵, S. A. Ward⁵ & S. Krishna¹

¹Department of Cellular and Molecular Medicine, St George's Hospital Medical School, Cranmer Terrace, London SW17 0RE, UK

²Department of Biochemistry and Molecular Biology, University of Southampton, Southampton SO16 7PX, UK

³Radioisotope Centre, Osaka City University Medical School, Asahi-machi 1-4-3, Abeno-ku, Osaka 545-8585, Japan

⁴Department of Chemistry, The Robert Robinson Laboratories, University of Liverpool, Liverpool L69 7ZD, UK

⁵Molecular and Biochemical Parasitology Group, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA, UK

Artemisinins are extracted from sweet wormwood (*Artemisia annua*) and are the most potent antimalarials available¹, rapidly killing all asexual stages of *Plasmodium falciparum*². Artemisinins are sesquiterpene lactones widely used to treat multidrug-resistant malaria¹, a disease that annually claims 1 million lives. Despite extensive clinical and laboratory experience^{3–5} their molecular target is not yet identified. Activated artemisinins form adducts with a variety of biological macromolecules, including haem, translationally controlled tumour protein (TCTP) and other higher-molecular-weight proteins⁶. Here we show that artemisinins, but not quinine or chloroquine, inhibit the SERCA orthologue (PfATP6) of *Plasmodium falciparum* in *Xenopus* oocytes with similar potency to thapsigargin (another sesquiterpene lactone and highly specific SERCA inhibitor). As predicted, thapsigargin also antagonizes the parasitocidal activity of artemisinin. Desoxyartemisinin lacks an endoperoxide bridge and is ineffective both as an inhibitor of PfATP6 and as an antimalarial. Chelation of iron by desferrioxamine abrogates the antiparasitic activity of artemisinins and correspondingly attenuates inhibition of PfATP6. Imaging of parasites with BODIPY-thapsigargin labels the cytosolic compartment and is competed by artemisinin. Fluorescent artemisinin labels parasites similarly and irreversibly in an Fe²⁺-dependent manner. These data provide compelling evidence that artemisinins act by inhibiting PfATP6 outside the food vacuole after activation by iron.

Some^{3,7}, but not all^{8–10}, studies suggest that artemisinins act by haem-dependent activation of an endoperoxide bridge occurring within the parasite's food vacuole. However, localization of artemisinins to parasite and not food vacuole membranes¹¹, and killing of tiny rings lacking haemozoin argue against the food vacuole being a major site for drug action². Artemisinins show structural similarities to thapsigargin, which is a highly specific inhibitor of sarco/endoplasmic reticulum Ca²⁺-ATPase (SERCA). We therefore hypothesized that when activated, artemisinins act by specifically and selectively inhibiting the SERCA of *P. falciparum*.

PfATP6 is the only SERCA-type Ca²⁺-ATPase sequence in the parasite's genome. We expressed PfATP6 in *Xenopus laevis* oocytes because it could not be functionally assayed in COS cells (J.M.E., unpublished work). Membrane preparations from oocytes expressing PfATP6 consistently gave higher Ca²⁺-dependent ATPase activities than water-injected control oocytes (mean $\pm$ s.e.m. activities of 0.054 $\pm$ 0.006 IU compared with 0.026 $\pm$ 0.003 international units (IU), n = 25; P < 0.0001). These activities are comparable to those observed with PfATP4, another *P. falciparum* Ca²⁺-ATPase¹². Figure 1a compares the Ca²⁺-activation profiles of PfATP6 and rabbit skeletal muscle SERCA1a (sarco/endoplasmic reticulum preparation) and shows that PfATP6 and SERCA1a are acti-

vated by similar concentrations of Ca²⁺ ions, [Ca²⁺]_{free} (Fig. 1a and ref. 12). PfATP6 (like SERCA1a) is inhibited by thapsigargin and cyclopiazonic acid, but not by ouabain, a specific Na⁺/K⁺-ATPase inhibitor (Fig. 1b). Vanadate, which is a more general inhibitor of P-type ATPases, also inhibits PfATP6.

We then assessed the inhibitory properties of chloroquine (a 4-aminoquinoline), quinine (a cinchona alkaloid) and artemisinin on transporter proteins encoded by *P. falciparum* and expressed in *Xenopus* oocytes, as well as their effects on SERCA1a. Supra-therapeutic concentrations of these antimalarials did not inhibit glucose transport by the falciparum hexose transporter (PfHT), confirming that none of these structurally diverse compounds

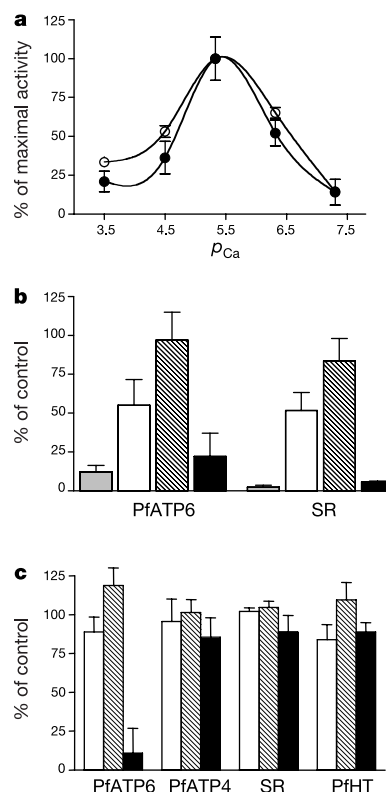


Figure 1 Functional characterization of PfATP6. **a**, Ca²⁺-dependency of PfATP6 activity (filled circles) compared with SERCA1a (SR) activity (open circles) is shown as a percentage of maximal activity (normalized to the maximum value, and not corrected for endogenous activity for PfATP6) at the indicated [Ca²⁺]_{free} values. $p_{Ca} = -\log_{10}[\text{Ca}^{2+}]_{\text{free}}$. Bars with each symbol represent standard errors for each assay (a minimum of three experiments for sarcoplasmic reticulum (SR) and seven for PfATP6). Half-maximal activation is achieved at [Ca²⁺]_{free} values of 0.64 μ M for PfATP6 and 0.3 μ M for SERCA. Maximal activation is at [Ca²⁺]_{free} of 48 μ M for both PfATP6 and SERCA. **b**, Inhibitor profiles for PfATP6 and SERCA. ATPase activities were determined with inhibitors and are compared with control values: thapsigargin (grey bar, 0.8 μ M, P < 0.0001; multivariate analysis of variance, MANOVA), sodium orthovanadate (white bar, 100 μ M, P = 0.043), ouabain (hatched bar, 100 μ M, P = 0.87), and cyclopiazonic acid (black bar, 1 μ M, P = 0.0033). Each bar represents a minimum of six experiments. **c**, Effects of antimalarials on two *P. falciparum* Ca²⁺-ATPases (PfATP6 and PfATP4), the *P. falciparum* hexose transporter (PfHT) and mammalian SERCA (SR). All activities are shown as a percentage of uninhibited activity and are compared with these activities: quinine (open bars, 10 μ M, P > 0.1; MANOVA), chloroquine (hatched bar, 1 μ M, P > 0.1), artemisinin (black bars; 50 μ M, P > 0.1 for PfATP4, SR, PfHT; 1 μ M for PfATP6, P < 0.0001). Ca²⁺-ATPase activity for PfATP6 is shown after correction of Ca²⁺-ATPase activity measured in water-injected oocytes, which was unnecessary for PfATP4, because there was no inhibition observed with any antimalarial. Each bar represents a minimum of four experiments.

interferes non-specifically with the function of integral membrane proteins of *P. falciparum* (Fig. 1c). SERCA1a activity was also unaffected by these antimalarials. Apart from PfATP6, PfATP4 is the only Ca^{2+} -ATPase-like sequence identified in the *P. falciparum* genome, and defines a non-SERCA subclass of Ca^{2+} -ATPase that is

unique to apicomplexan organisms¹². PfATP4 activity was also not inhibited by any of the antimalarials tested. In contrast, Ca^{2+} -dependent ATPase activity of PfATP6 was abolished by artemisinin, but not by quinine or chloroquine. These data show that even at relatively high concentrations (1 μM) artemisinin inhibits PfATP6

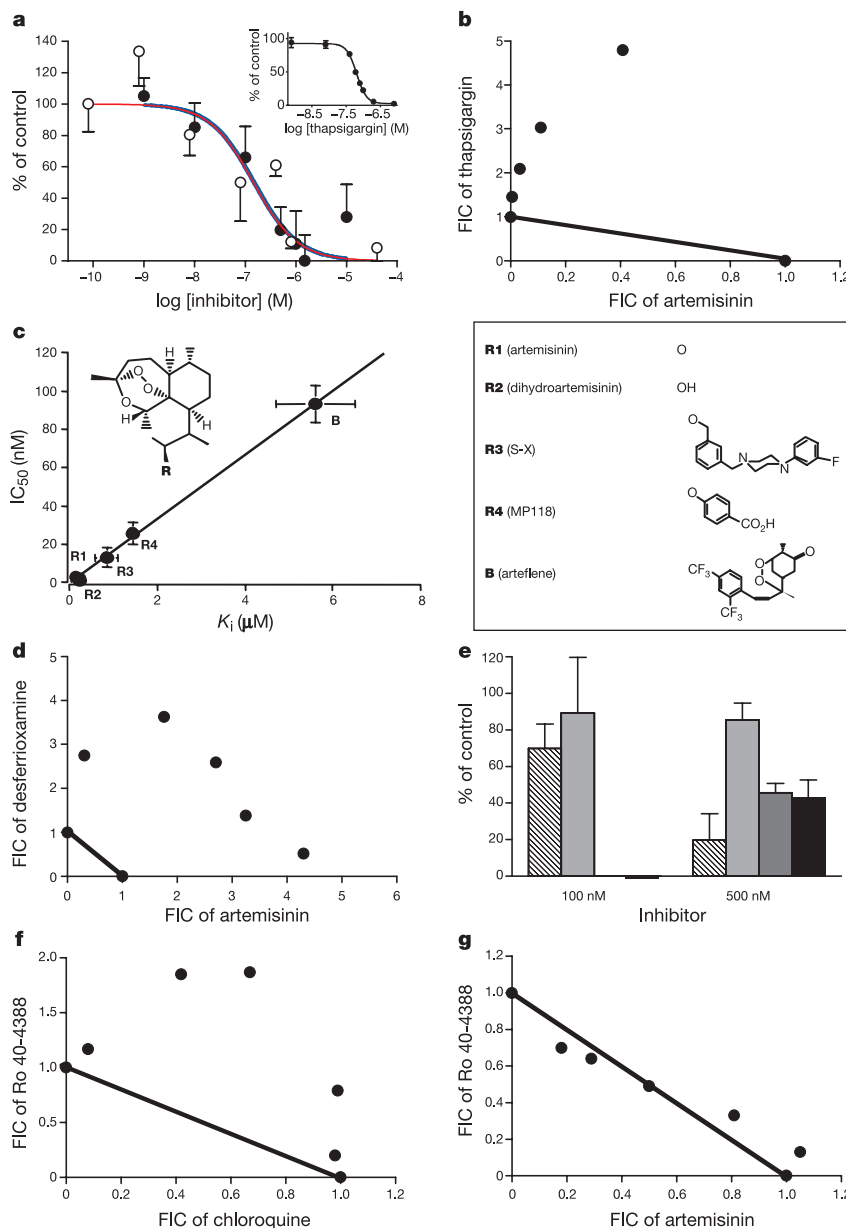


Figure 2 Properties of artemisinin inhibition. **a**, The half-maximal inhibition constants (K_i) for artemisinin (black circles, blue line, 162 ± 31 nM) and thapsigargin (white circles, red line, 146 ± 66 nM) were determined. For thapsigargin endogenous Ca^{2+} -ATPase activity was not subtracted, because K_i values for both activities are indistinguishable, and endogenous effects can be neglected. Because of the insensitivity of endogenous activity to artemisinin, this endogenous Ca^{2+} -ATPase component was determined at high artemisinin concentration (50 μM) in PfATP6-expressing oocytes, and used to correct values. K_i values for thapsigargin and artemisinin are not significantly different ($P = 0.58$). Data are from a minimum of three experiments. The inset shows the determination of K_i for SERCA1a-rich rabbit-muscle preparation ($K_i = 64$ nM). **b**, An isobologram for thapsigargin and artemisinin. The solid line indicates an isobole where drugs act additively and independently. Data points above this line indicate antagonism between drugs, and data below this line would indicate synergy³⁰. FIC is fractional inhibitory concentration. **c**, The relationship between inhibition of PfATP6 activities

determined in *Xenopus* oocyte preparations is shown compared with inhibitory potencies of artemisinins whose structures are also given. Data are mean values (from a minimum of three independent experiments) with standard error bars, where visible. **d**, An isobologram for desferrioxamine and artemisinin. Interpretation is as for **b**. **e**, PfATP6 activity (corrected for endogenous activity) was determined at two different artemisinin concentrations without (hatched bars) and in presence of desferrioxamine (100 μM , light grey bars). At 500 nM there is significant ($P = 0.0093$) abrogation of inhibition of PfATP6 by artemisinin. PfATP6 activity (not corrected for endogenous) was also determined at one thapsigargin concentration without (dark grey bars) and in the presence of desferrioxamine (100 μM , black bar). There is no significant difference in inhibitory potency ($P > 0.5$). Each bar represents at least six experiments. **f**, An isobologram for chloroquine and Ro 40-4388. Interpretation is as for **b**. **g**, An isobologram for artemisinin and Ro 40-4388. Interpretation is as for **b**.

with high specificity. No other transporter, including SERCA1a, is inhibited by artemisinin even when used at 50 times (50 μM , Fig. 1c) the concentration used to inhibit PfATP6 activity in this experiment (1 μM). Endogenous (oocyte) Ca^{2+} -dependent ATPase activity is also resistant to the action of artemisinin (data not shown), demonstrating the unique specificity of artemisinins for PfATP6.

To compare the interactions of artemisinin and thapsigargin with PfATP6, we measured inhibitory constants for these compounds in oocytes. Thapsigargin irreversibly inhibits Ca^{2+} -dependent ATPase activity of SERCA with 1:1 stoichiometry¹³. Remarkably, the inhibitory profile of artemisinin is exactly superimposable over that of thapsigargin (Fig. 2a), showing that they inhibit PfATP6 with the same potency and stoichiometry *in vitro*. Artemisinin concentrations that inhibit PfATP6 activity in oocytes ($K_i \approx 150 \text{ nM}$) are somewhat higher than those required to kill cultured parasites (median $\text{IC}_{50} = 5.47 \text{ nM}$ (range $\text{IC}_{50} = 0.47 - 27.1 \text{ nM}$) artesunate; $n = 256$ (ref. 14), and Fig. 2c).

This is unsurprising for several reasons. First, there is no information on the concentrations of artemisinins at their sub-cellular sites of action, although artemisinins are clearly concentrated in infected erythrocytes by several hundred-fold¹⁵. Second, the extent of inhibition of PfATP6 required to kill parasites is not yet determined (it may, for example, be less than the K_i value determined in oocytes). Third, our *in vitro* assays are conducted over relatively short periods of time (<15 min) whereas parasite killing is determined after exposure to drug for many hours, and killing is also enhanced in a time-dependent manner². Fourth, the oocyte membrane preparation that we use has relatively little target protein (about 1%¹²) in relation to the amount of lipid, favouring sequestration of lipid-soluble components away from target protein.

Figure 2a (inset) shows that the K_i for thapsigargin on SERCA1a-rich rabbit-muscle preparations (containing protein 80% of which is SERCA1a) is 64 nM, almost double the IC_{50} value for lymphocytes (35 nM).

As both artemisinin and thapsigargin specifically inhibit PfATP6 after heterologous expression, we predicted that they would exhibit mutual antagonism in cultured parasites. Artemisinin and thapsigargin were administered simultaneously in varying concentrations, and the results plotted as an isobologram (Fig. 2b). The plot shows significant antagonism of the two drugs, indicating that artemisinin specifically inhibits PfATP6 in intact parasites, just as it does in the heterologous expression system. The higher IC_{50} value for thapsigargin in parasite culture ($\text{IC}_{50} = 2.55 \mu\text{M}$) compared with the inhibitory constant for PfATP6 in oocyte membranes ($K_i = 146 \text{ nM}$) may reflect difficulties in accessing intraparasitic PfATP6. This is probably because multiple membranes need to be crossed in parasites, but not in mammalian cells, which are killed by nanomolar concentrations of thapsigargin (discussed below).

To confirm that PfATP6 is the target for artemisinins we compared inhibitory constants of a range of artemisinin derivatives against PfATP6 in oocytes, with their ability to kill parasites *in vitro*. Figure 2c shows results for artemisinin, dihydroartemisinin (the principal antimalarial metabolite of many artemisinin derivatives including artesunate, artemether and artemotil), two other artemisinin derivatives (MP118 and S-X) and arteflene, a less effective semi-synthetic endoperoxide derivative. There is almost perfect correlation between inhibitory potencies of artemisinins for PfATP6 activity and IC_{50} values from cultured parasites ($r^2 = 0.999$, Fig. 2c). Those artemisinins that are active antimalarials all contain an endoperoxide bridge¹⁶, a key structural feature

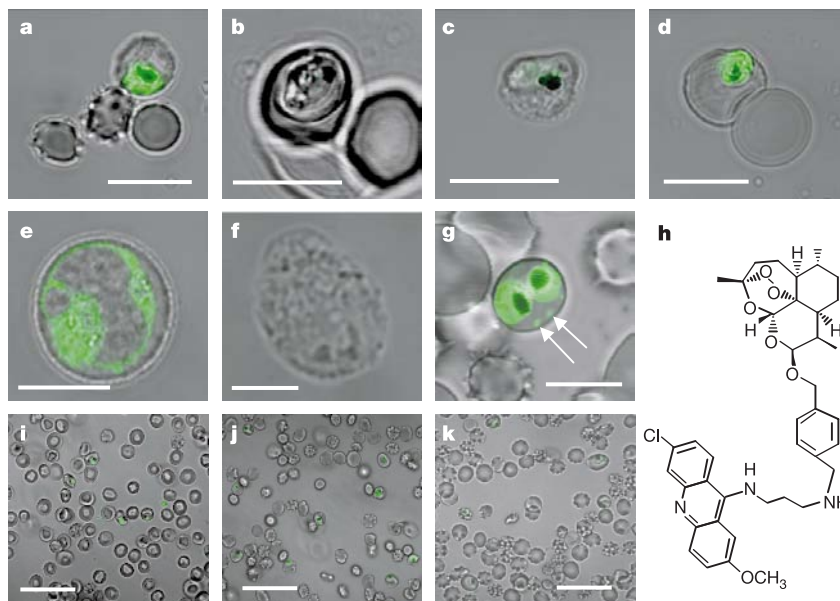


Figure 3 *In vivo* labelling with BODIPY-thapsigargin and fluorescently-labelled artemisinin. Cells were incubated in 200 nM BODIPY-thapsigargin (10 min, 37 °C, in presence of 0.025% Pluronic F127), then washed three times and imaged. Pictures were taken after a preliminary scan, to minimize bleaching. Preincubations with unlabelled substances were performed 40 min (37 °C) before incubation with BODIPY-thapsigargin. Artemisinin labelling (500 nM) was performed in absence of Pluronic F127. Scale bars represent 10 μm , unless stated otherwise. **a**, Fluorescence after incubation in BODIPY-thapsigargin in a single infected cell. **b**, BODIPY-thapsigargin signal detected after preincubation in thapsigargin (50 μM). **c**, BODIPY-thapsigargin signal detected after preincubation in artemisinin (50 μM). **d**, BODIPY-thapsigargin signal detected after preincubation in artemisinin (50 μM) and desferrioxamine (100 μM). **e**, BODIPY-thapsigargin distribution

in the lymphocyte cell line PM-1, showing typical SERCA-type distribution. **f**, PM-1 cell labelled with BODIPY-thapsigargin after competition with unlabelled thapsigargin (50 μM). **g**, Distribution of labelled artemisinin in live parasites. Arrows indicate artemisinin in erythrocyte cytosol. **h**, Chemical structure of labelled artemisinin. **i**, Low magnification image of live parasites labelled with artemisinin (scale bar represents approximately 30 μm). **j**, After incubation in labelled artemisinin, cells were washed in 10-fold excess of unlabelled artemisinin and imaged (scale bar represents approximately 30 μm). **k**, Before incubation in labelled artemisinin, cells were pre-treated with desferrioxamine (100 μM , 30 min), and after labelling washed as in **j** (scale bar represents approximately 30 μm).

that generates carbon-centred short-lived radicals through iron-mediated catalysis. These intermediate radicals react with and incapacitate target protein(s). Desoxyderivatives are otherwise structurally identical to artemisinins but lack both an endoperoxide bridge and significant antimalarial activity. We predicted that desoxyartemisinin would be a significantly less effective inhibitor of PfATP6. Desoxyartemisinin does not inhibit PfATP6 activity even at concentrations that are two orders of magnitude greater (50 μM) than inhibitory concentrations of artemisinin itself. Again, this observation is consistent with the IC_{50} value for desoxyartemisinin determined against parasites *in vitro* (mean $4.85 \pm 1.7 \mu\text{M}$).

The generation of biologically reactive species from artemisinins depends on cleavage of the peroxide bridge via an iron-dependent mechanism. Consequently, in cultures of *P. falciparum*, iron deprivation using tight-binding chelators such as desferrioxamine may inhibit antimalarial activity¹⁷. We confirmed this antagonistic interaction between iron-deprivation and artemisinins¹⁷ by isobologram analysis (Fig. 2d). If our hypothesis is correct, iron chelation should also specifically block the inhibitory effects of artemisinin in our oocyte model. Desferrioxamine abolishes the inhibitory activity of artemisinin on PfATP6 but does not alter the inhibitory properties of thapsigargin (Fig. 2e). On the basis of these observations and published data, we suggest the following model for the antimalarial action of artemisinins.

Artemisinins produce carbon-centred free radicals in the presence of catalytic quantities of Fe^{2+} with the selective targeting of PfATP6 by activated artemisinins being determined by structural features that depend on their sesquiterpene moiety. This proposed mechanism of action contrasts with suggestions that artemisinins may, like chloroquine, bind to haem and interfere with the process of haem crystallization. We therefore tested the relevance of this latter hypothesis using a protease inhibitor (Ro 40-4388)¹⁸ that inhibits plasmepsin activity, blocks the first step in haemoglobin degradation and prevents the release of haem *in situ*. Chloroquine and Ro 40-4388 exhibit mutual antagonism in isobologram analysis (Fig. 2f) as expected if the mode of action of chloroquine stems from binding of the drug to haem, whereas there is neither antagonism nor synergism between Ro 40-4388 and artemisinin (Fig. 2g). This finding confirms that Fe^{2+} -dependent activation and antimalarial activity of artemisinins is independent of haem binding.

To prove a direct interaction between artemisinins and PfATP6, we used a fluorescent derivative of thapsigargin (BODIPY-thapsigargin) to localize it in parasites (Fig. 3a) and demonstrated that competition with an excess of unlabelled thapsigargin abolished this characteristic labelling of parasites (Fig. 3b; for example from 3.7% labelled parasite cells to 0%). Competition was also observed with an excess of artemisinin (Fig. 3c; reducing labelling from 3.7% to 0.84%; $P < 0.0001$; in one example of three independent experiments), confirming that both artemisinin and thapsigargin interact at a similar site within the parasite. Consistent with the requirement for iron to activate artemisinin, adding desferrioxamine together with artemisinin abolished competition for thapsigargin-binding sites and restored labelling of parasites by BODIPY-thapsigargin (Fig. 3d). To confirm that interaction with PfATP6 was specific for artemisinin, we examined labelling of parasites by thapsigargin in the presence of quinine and found that this labelling was unaffected (not shown). We further confirmed the specificity of these results using BODIPY-thapsigargin to demonstrate the characteristic SERCA pattern of labelling in mammalian cells (PM-1, Fig. 3e)¹⁹ and abolition of labelling after preincubation with an excess of unlabelled thapsigargin (Fig. 3f).

We next synthesized a fluorescent artemisinin derivative (Fig. 3h) and used this to image the distribution of artemisinin in infected erythrocytes (Fig. 3g). Artemisinin distributes like BODIPY-thapsigargin to membranous structures in the cytoplasm of parasites, and does not localize to the parasite food vacuole. Unlike BODIPY-thapsigargin, the fluorescent artemisinin derivative also

labels membranous structures within the cytoplasm of the host erythrocyte (Fig. 3g, arrows); these structures are identical to a tubovesicular membranous network that transports artemisinin through the host cell compartment and into the intraerythrocytic parasite²⁰. Artemisinins are thought to interact initially with a carrier before accessing the tubovesicular membranous network, resulting in a highly efficient delivery pathway²⁰. The fact that similar erythrocytic structures are not observed with BODIPY-thapsigargin suggests that a comparable mechanism of delivery to parasites does not exist. Poor penetration of infected erythrocytes by thapsigargin may explain why cultured parasites are relatively insensitive to the compound, even though inhibitory constants for thapsigargin and artemisinin against PfATP6 expressed in oocytes are similar. We further show in Fig. 3i–k that *in vivo* binding of artemisinin to parasites is not diminished by an excess of unlabelled artemisinin, unless parasites have been previously exposed to desferrioxamine, when a substantial proportion of labelled artemisinin is removable.

Artemisinin forms covalent adducts with four major membrane-associated parasite proteins (relative molecular mass, M_r , 25K, 50K, 65K and $>200\text{K}$) as shown in previous studies²¹. Only one of these adducts has been investigated (TCTP, 25K)²², and it has not been functionally characterized, so that inhibition of other parasite proteins by artemisinins may still play an as-yet-unspecified role. A dimeric configuration for PfATP6 would predict an adduct of about 280K, consistent with the largest molecular species labelled by artemisinin. Furthermore, exposure to artemisinins rapidly causes swelling of endoplasmic reticulum in parasites, providing a morphological correlate for our proposed site of action²³.

Withering²⁴ first described the therapeutic use of purple foxglove extracts containing digitalis in 1785, although mammalian Na^+/K^+ -ATPases were not identified as the drug target of digitoxin, an active compound in these preparations, until 1953 (ref. 25). Our studies suggest that *Artemisia* extracts, used for an even longer period of time in China, also selectively inhibit a P-type ATPase of the malarial parasite. These results validate P-type ATPases as a drug target of human pathogens. The recent structural solution of mammalian SERCAs will aid in understanding selective toxicity of artemisinins and in producing congeners^{13,26}. □

Methods

Expression studies

For functional studies in *X. laevis* oocytes the gene encoding PfATP6 was ligated into pMos blue (linearized with EcoRV) using a TA cloning strategy after strengthening of the Kozak consensus (accession number AJ532679). It was then sub-cloned into the *Xenopus* expression vector pcDNA3.1 (+) using flanking BamHI and XbaI restriction sites, and linearized with XbaI. Capped complementary RNA was transcribed (T7-mMESSAGE mMACHINE kit, Ambion, Austin, TX)²⁷. *X. laevis* oocytes were prepared, microinjected, and total membrane preparations harvested after 4–5 days as described^{12,27}. Sarcoplasmic reticulum (enriched for SERCA1a) was prepared from skeletal muscle of New Zealand White rabbits as described²⁸.

Ca^{2+} -ATPase assay

Ca^{2+} -ATPase activity was measured (10 μg total protein, 25 °C) using a coupled enzyme assay as described²⁸ (CamSpec, Cambridge, UK). $[\text{Ca}^{2+}]_{\text{free}}$ was calculated using MaxChelator²⁹.

Glucose uptake by PfHT

Uptake of permeant D-[U-¹⁴C]glucose was measured as described²⁷.

Parasites and parasite culture

Parasites (clones 3D7, HB3 and K1) were cultured and inhibitory constants for drugs assayed as described². The effect of the interaction between two antimalarial agents on the growth of *P. falciparum* in culture was assessed using isobole analysis, as described³⁰.

Confocal microscopy

Cells (at room temperature) were imaged after labelling with BODIPY-thapsigargin (B-7487, Molecular Probes) using a LSM 510 scanning confocal microscope (Carl Zeiss, Jena, Germany), with excitation at 488 nm attenuated to 0.2–0.5% intensity with an acousto-optical tunable filter. Pinhole size was adjusted to give optical sections $<1.2 \mu\text{m}$, emission was detected at $>505 \text{ nm}$ and analysed with LSM 510 software (v.2.01, Carl Zeiss). For imaging of fluorescent artemisinin, medium-term trophozoite-infected

erythrocytes were immobilized using poly-L-lysine coated coverslips in a Bioprocess FCS2 perfusion chamber (maintained at 37 °C in normal growth medium). Fluorescent-labelled artemisinin (500 nM) was added 10 min before imaging.

Received 20 March; accepted 12 May 2003; doi:10.1038/nature01813.

- Hien, T. T. & White, N. J. Qinghaosu. *Lancet* **341**, 603–608 (1993).
- ter Kuile, F., White, N. J., Holloway, P. H., Pasvol, G. & Krishna, S. *Plasmodium falciparum*: In vitro studies of the pharmacodynamic properties of drugs used for the treatment of severe malaria. *Exp. Parasitol.* **76**, 85–95 (1993).
- Jefford, C. W. Why artemisinin and certain synthetic peroxides are potent antimalarials. Implications for the mode of action. *Curr. Med. Chem.* **8**, 1803–1826 (2001).
- Robert, A., Dechy-Cabaret, O., Cazes, J. & Meunier, B. From mechanistic studies on artemisinin derivatives to new modular antimalarial drugs. *Acc. Chem. Res.* **35**, 167–174 (2002).
- Olliaro, P. L., Haynes, R. K., Meunier, B. & Yuthavong, Y. Possible modes of action of the artemisinin-type compounds. *Trends Parasitol.* **17**, 122–126 (2001).
- Meshnick, S. R. Artemisinin: mechanisms of action, resistance and toxicity. *Int. J. Parasitol.* **32**, 1655–1660 (2002).
- Pandey, A. V., Tekwani, B. L., Singh, R. L. & Chauhan, V. S. Artemisinin, an endoperoxide antimalarial, disrupts the hemoglobin catabolism and heme detoxification systems in malarial parasite. *J. Biol. Chem.* **274**, 19383–19388 (1999).
- Haynes, R. K. *et al.* Artemisinin antimalarials do not inhibit hemozoin formation. *Antimicrob. Agents Chemother.* **47**, 1175 (2003).
- O'Neill, P. *et al.* Biomimetic Fe(II)-mediated degradation of arteflene (Ro-42-1611). The first EPR spin-trapping evidence for the previously postulated secondary carbon-centered cyclohexyl radical. *J. Org. Chem.* **65**, 1578–1582 (2000).
- Hawley, S. R. *et al.* Relationship between antimalarial drug activity, accumulation, and inhibition of heme polymerization in *Plasmodium falciparum* in vitro. *Antimicrob. Agents Chemother.* **42**, 682–686 (1998).
- Ellis, D. S. *et al.* The chemotherapy of rodent malaria, XXXIX. Ultrastructural changes following treatment with artemisinin of *Plasmodium berghei* infection in mice, with observations of the localization of [³H]-dihydroartemisinin in *P. falciparum* in vitro. *Ann. Trop. Med. Parasitol.* **79**, 367–374 (1985).
- Krishna, S. *et al.* Expression and functional characterization of a *Plasmodium falciparum* Ca²⁺-ATPase (PfATP4) belonging to a subclass unique to apicomplexan organisms. *J. Biol. Chem.* **276**, 10782–10787 (2001).
- Toyoshima, C. & Nomura, H. Structural changes in the calcium pump accompanying the dissociation of calcium. *Nature* **418**, 605–611 (2002).
- Price, R. *et al.* The *pfmdr1* gene is associated with a multidrug-resistant phenotype in *Plasmodium falciparum* from the western border of Thailand. *Antimicrob. Agents Chemother.* **43**, 2943–2949 (1999).
- Gu, H. M., Warhurst, D. C. & Peters, W. Uptake of [³H] dihydroartemisinin by erythrocytes infected with *Plasmodium falciparum* in vitro. *Trans. R. Soc. Trop. Med. Hyg.* **78**, 265–270 (1984).
- Haynes, R. K. Artemisinin and derivatives: the future for malaria treatment? *Curr. Opin. Infect. Dis.* **14**, 719–726 (2001).
- Meshnick, S. R. *et al.* Iron-dependent free radical generation from the antimalarial agent artemisinin (qinghaosu). *Antimicrob. Agents Chemother.* **37**, 1108–1114 (1993).
- Bray, P. G., Munghin, M., Ridley, R. G. & Ward, S. A. Access to hemozoin: the basis of chloroquine resistance. *Mol. Pharmacol.* **54**, 170–179 (1998).
- Simpson, P. B. & Russell, J. T. Role of sarcoplasmic/endoplasmic-reticulum Ca²⁺-ATPases in mediating Ca²⁺ waves and local Ca²⁺-release microdomains in cultured glia. *Biochem. J.* **325**, 239–247 (1997).
- Akompong, T., VanWye, J., Ghorri, N. & Haldar, K. Artemisinin and its derivatives are transported by a vacuolar-network of *Plasmodium falciparum* and their anti-malarial activities are additive with toxic sphingolipid analogues that block the network. *Mol. Biochem. Parasitol.* **101**, 71–79 (1999).
- Asawamahakada, W., Ittarat, I., Pu, Y. M., Ziffer, H. & Meshnick, S. R. Reaction of antimalarial endoperoxides with specific parasite proteins. *Antimicrob. Agents Chemother.* **38**, 1854–1858 (1994).
- Bhisuttibhhan, J. *et al.* The *Plasmodium falciparum* translationally controlled tumor protein homolog and its reaction with the antimalarial drug artemisinin. *J. Biol. Chem.* **273**, 16192–16198 (1998).
- Ono, T. *et al.* Degenerative changes in morphology of *Plasmodium falciparum* induced by artemether in vitro. *Jpn J. Parasitol.* **40**, 587–595 (1991).
- Allen, D. G., Eisner, D. A. & Wray, S. C. Birthday present for digitalis. *Nature* **316**, 674–675 (1985).
- Schatzmann, H.-J. Herzglykoside als Hemmstoffe für den aktiven Kalium- und Natriumtransport durch die Erythrocytenmembran. *Helv. Physiol. Acta* **11**, 346–354 (1953).
- Toyoshima, C., Nakasako, M., Nomura, H. & Ogawa, H. Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution. *Nature* **405**, 647–655 (2000).
- Woodrow, C. J., Penny, J. I. & Krishna, S. Intraerythrocytic *Plasmodium falciparum* expresses a high-affinity facilitative hexose transporter. *J. Biol. Chem.* **274**, 7272–7277 (1999).
- East, J. M. Purification of a membrane protein (Ca²⁺/Mg²⁺-ATPase) and its reconstitution into lipid vesicles. *Methods Mol. Biol.* **27**, 87–94 (1994).
- Bers, D. M., Patton, C. W. & Nuccitelli, R. A practical guide to the preparation of Ca²⁺ buffers. *Methods Cell Biol.* **40**, 3–29 (1994).
- Berenbaum, M. C. A method for testing for synergy with any number of agents. *J. Infect. Dis.* **137**, 122–130 (1978).

Acknowledgements We thank T. Joët and A.-C. Uhlemann for discussions, P. Wünnenberg for technical assistance, A. Craig for a *P. falciparum* cDNA library and K. Tanabe for the *P. falciparum* genomic clone 3L6. We thank the charity Hope (Wessex Medical Trust) for financial support of J.M. East, and T. Bolton for access to confocal microscopy. U. E.-L. was funded by the Deutsche Forschungsgemeinschaft, S.A.W. and P.G.B. are funded by the Wellcome Trust and S.K. holds an MRC (UK) grant. This paper is dedicated to the memory of G. Cowan.

Competing interests statement The authors declare competing financial interests: details accompany the paper on www.nature.com/nature.

Correspondence and requests for materials should be addressed to S.K. (s.krishna@sghms.ac.uk).

Feedback regulation of MAPK signalling by an RNA-binding protein

Reiko Sugiura¹, Ayako Kita¹, Yasuhito Shimizu¹, Hisato Shuntoh², Susie O. Sio^{1*} & Takayoshi Kuno¹

¹Division of Molecular Pharmacology and Pharmacogenomics, Department of Genome Sciences, Kobe University Graduate School of Medicine, Kobe 650-0017, Japan

²Faculty of Health Science, Kobe University School of Medicine, 7-10-2 Tomogaoka, Suma-ku, Kobe 654-0142, Japan

* Present address: Department of Pharmacology, College of Medicine, University of the Philippines Manila, Manila 1000, Philippines

Mitogen-activated protein kinases (MAPKs) are evolutionarily conserved enzymes that convert extracellular signals into various outputs such as cell growth, differentiation and cell death^{1–4}. MAPK phosphatases selectively inactivate MAPKs by dephosphorylating critical phosphothreonine and phosphotyrosine residues^{5,6}. The transcriptional induction of MAPK phosphatase expression by various stimuli, including MAPK activation, has been well documented as a negative-feedback mechanism of MAPK signalling^{7,8}. Here we show that Rnc1, a novel K-homology-type RNA-binding protein in fission yeast, binds and stabilizes Pmp1 messenger RNA⁹, the MAPK phosphatase for Pmk1 (refs 10, 11). Rnc1 therefore acts as a negative regulator of Pmk1 signalling. Notably, Pmk1 phosphorylates Rnc1, causing enhancement of the RNA-binding activity of Rnc1. Thus, Rnc1 is a component of a new negative-feedback loop that regulates the

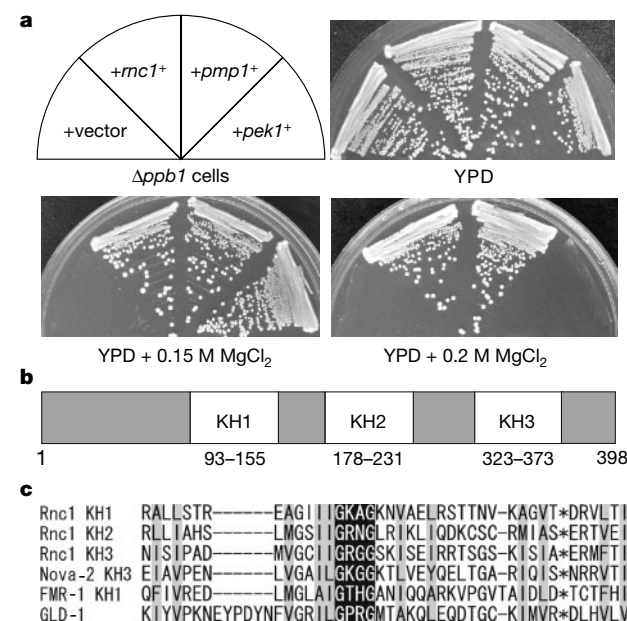


Figure 1 Identification of Rnc1, a new KH-type RNA-binding protein. **a**, *mrc1*⁺ overexpression suppresses the Cl[−] sensitivity of calcineurin deletion (*Δppb1*). Cells were transformed with a multicopy plasmid containing the indicated genes, and streaked onto each plate, then incubated at 30 °C. **b**, Schematic overview of Rnc1. The regions containing the KH motifs are represented by open boxes. Amino acid positions are indicated. **c**, KH-domain sequence alignment in the three KH domains of Rnc1, human Nova-2 KH3, human FMR-1 KH1 and *Caenorhabditis elegans* GLD-1. Black shading denotes the invariant Gly-X-X-Gly segment; grey shading indicates the aliphatic α/β platform; asterisk indicates the variable loop¹⁹.

Pmk1 pathway through its binding to Pmp1 mRNA. Our findings—the post-transcriptional mRNA stabilization of a MAPK phosphatase mediated by an RNA-binding protein—provide an additional regulatory mechanism for fine-tuning of MAPK signalling pathways.

Disruption of the calcineurin gene *ppb1*⁺ (ref. 12) in the fission yeast *Schizosaccharomyces pombe* results in severe sensitivity to Cl[−] (Fig. 1a, +vector)⁹. We performed a genetic screen with the aim of identifying genes that when overexpressed could suppress the Cl[−]-sensitive growth defect of calcineurin deletion⁹ (*Δppb1*). Using this screen we previously identified two regulators of the Pmk1 pathway¹⁰: the *pmp1*⁺ gene (Fig. 1a, +*pmp1*⁺), which encodes a MAPK phosphatase that inactivates Pmk1⁹; and the *pek1*⁺ gene (Fig. 1a, +*pek1*⁺), which encodes a MAPK kinase (MAPKK) for Pmk1, and which has been shown to inhibit Pmk1 signalling in its unphosphorylated state¹³.

To identify new regulators of Pmk1 signalling, we further screened and identified a gene that when overexpressed could

suppress the Cl[−] sensitivity of *Δppb1* (Fig. 1a). The nucleotide sequence revealed an open reading frame (SPCC757.09c) encoding a protein containing three K-homology (KH) motifs (Fig. 1b). The KH motifs, originally identified in hnRNP-K¹⁴, are almost completely evolutionarily conserved and have been implicated in RNA binding^{15,16} (Fig. 1c). We therefore named this gene *rnc1*⁺ (for RNA-binding protein that suppresses calcineurin deletion 1).

The above findings led to the hypothesis that Rnc1 is functionally related to Pmk1 signalling. Thus, we examined the effects of deletion and overexpression of *rnc1*⁺ on Pmk1 signalling *in vivo*. Phosphorylation of Pmk1 was more heavily detected in *rnc1*-deleted cells (*Δrnc1*) compared with wild-type cells both before and after heat shock (48 °C for 20 min)¹⁷ (Fig. 2a). In contrast, no Pmk1 phosphorylation was detected in wild-type cells overproducing Rnc1 (Rnc1 OP) (Fig. 2a). Thus, Rnc1 negatively regulates the degree of phosphorylation of Pmk1.

This raises the possibility that Rnc1 affects the metabolism of specific cellular mRNAs involved in the regulation of Pmk1 signalling. Therefore we examined the effects of *rnc1* deletion on the mRNA level of the MAPK phosphatase Pmp1, a candidate target for Rnc1. We examined the mRNA level of glutathione S-transferase (GST)–Pmp1 transcribed from the thiamine-repressible *nmt1* promoter¹⁸. Levels of GST–Pmp1 mRNA were significantly lower in *Δrnc1* cells compared with wild-type cells (Fig. 2b, 0 h, left panel). Furthermore, the GST–Pmp1 mRNA level in *Δrnc1* cells decreased more rapidly after shutting off transcription, and mRNA was undetectable 6 h after addition of thiamine (Fig. 2b). Consistently, *pmp1*⁺ overexpression failed to suppress the Cl[−] sensitivity of

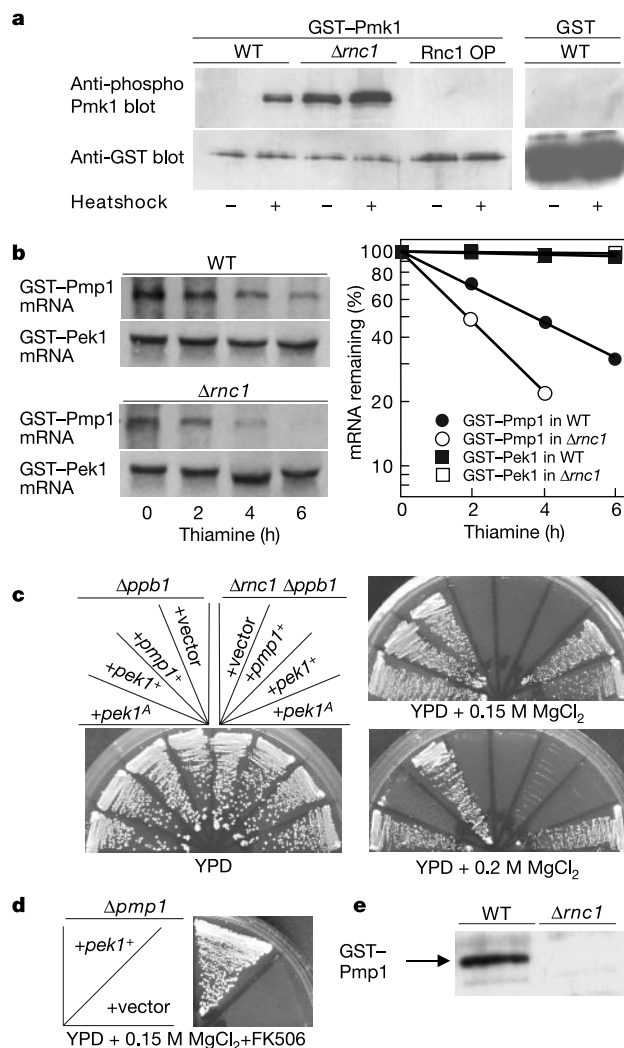


Figure 2 Rnc1 modulates MAPK signalling through regulation of the MAPK phosphatase Pmp1. **a**, Rnc1 regulates the phosphorylation level of Pmk1 *in vivo*. Phosphorylation of Pmk1 was examined using anti-phospho Pmk1 antibodies. **b**, Pmp1 mRNA is destabilized by *Δrnc1* mutation. The stability of GST–Pmp1 and GST–Pek1 mRNAs was analysed as described in Methods. **c**, *pmp1*⁺ overexpression failed to suppress the Cl[−] sensitivity of calcineurin deletion in *Δrnc1* cells. **d**, *pek1*⁺ overexpression suppressed the Cl[−] sensitivity of *Δpmp1* mutants. **e**, Immunoblot analysis of Pmp1 protein levels in wild-type and *Δrnc1* cells. GST–Pmp1 was probed with anti-GST antibodies.

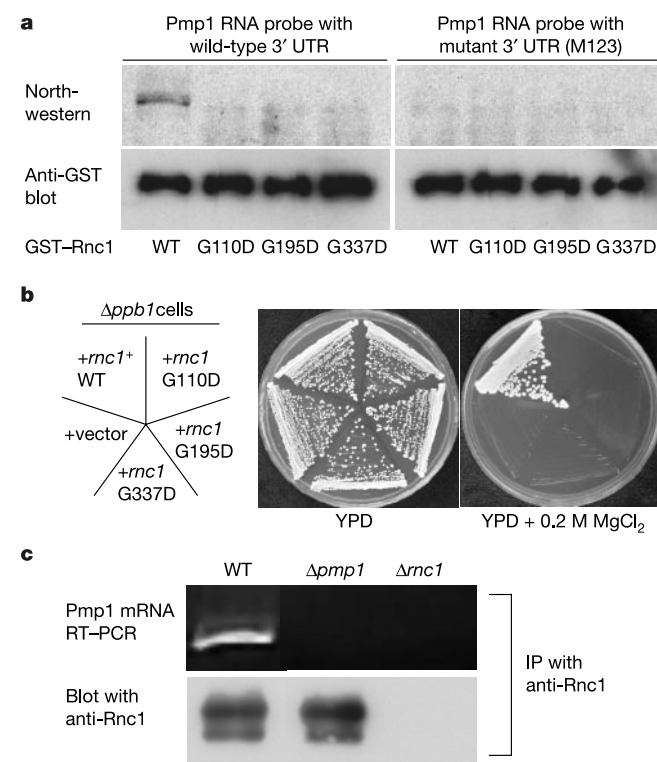


Figure 3 Rnc1 binds Pmp1 mRNA *in vitro* and *in vivo*. **a**, Northwestern blot analysis of Rnc1 and its KH-domain mutants. Purified GST-fusion proteins were blotted and probed with Pmp1 mRNA with wild-type or mutant 3' UTR (M123, see Supplementary Fig. S1a). **b**, Rnc1 KH-domain mutants fail to suppress the Cl[−] sensitivity of calcineurin deletion. *Δppb1* cells were transformed with a multicopy plasmid containing the indicated genes, and were analysed as in Fig. 1a. **c**, Rnc1 associates with Pmp1 mRNA *in vivo*. Wild-type, *Δpmp1* or *Δrnc1* cells were analysed by immunoprecipitation and RT–PCR as described in Methods.

$\Delta rnc1$ $\Delta ppb1$ double mutants, whereas it fully suppressed the Cl^- sensitivity of $\Delta ppb1$ mutants (Fig. 2c). Levels of GST-Pek1 mRNA, in contrast, were stable and were not affected by $rnc1$ deletion (Fig. 2b). Although the $\Delta rnc1$ $\Delta ppb1$ double mutant was viable, it showed more severe Cl^- sensitivity than did $\Delta ppb1$ cells (data not shown). Overproduction of wild-type Pek1 suppressed the Cl^- sensitivity of both $\Delta rnc1$ $\Delta ppb1$ and $\Delta ppb1$ mutants (Fig. 2c), and the unphosphorylatable form of Pek1—Pek1^A (Pek1(T238A))—showed greater suppression of Cl^- sensitivity compared with wild-type Pek1 (Fig. 2c). This is consistent with our previous study, which showed that Pek1, in its unphosphorylated form, acts as a potent negative regulator of Pmk1 signalling¹³. Furthermore, although $\Delta pmp1$ and $\Delta ppb1$ are synthetically lethal⁹, the overexpression of wild-type Pek1 suppressed the Cl^- sensitivity of $\Delta pmp1$ cells in the presence of the specific calcineurin inhibitor FK506 (Fig. 2d). Together, these results indicate that Pek1 overexpression still downregulates Pmk1 even in the absence of Rnc1 or Pmp1, and they support the hypothesis that Pmp1, but not

Pek1, is a target of Rnc1. Consistently, the protein level of GST-Pmp1 in $\Delta rnc1$ cells was barely detectable (Fig. 2e).

To test whether Rnc1 directly binds Pmp1 mRNA, we performed northwestern blot analysis with purified GST-Rnc1 protein using Pmp1 mRNA as a probe. With wild-type Rnc1, the binding signal in northwestern blots corresponded to the prominent protein band on the western blot, indicating that GST-fused wild-type Rnc1 bound Pmp1 RNA (Fig. 3a, left panel, WT). We then examined the effect of the Gly to Asp point mutation in the first glycine of Gly-X-X-Gly, a highly conserved consensus sequence that is important for RNA binding¹⁹ and which is located in each of the three KH domains in Rnc1 (Fig. 1c). In contrast to wild-type Rnc1, all three KH-domain mutants showed a barely detectable binding to Pmp1 RNA (Fig. 3a, left panel). Consistently, the expression of these three KH-domain mutants all failed to suppress the Cl^- sensitivity of calcineurin deletion (Fig. 3b). Together, these observations indicate that Rnc1 directly binds Pmp1 mRNA, and that this interaction is mediated, at least in part, by the KH motif.

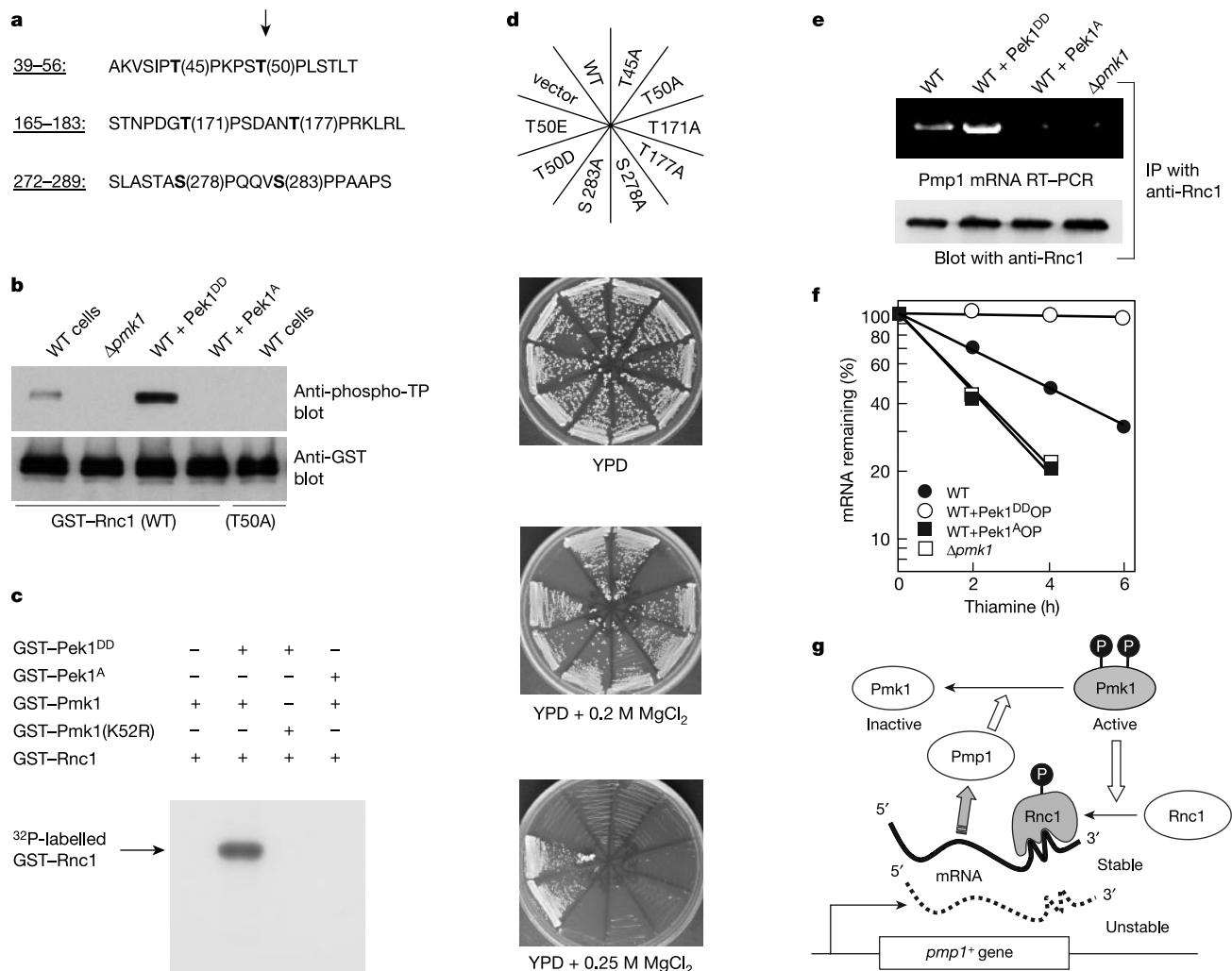


Figure 4 Pmk1-mediated phosphorylation regulates the ability of Rnc1 to bind and stabilize Pmp1 mRNA. **a**, Amino acid sequence containing the six putative MAPK phosphorylation sites. The identified phosphorylation site (T50) is noted with an arrow. **b**, Pmk1 regulates the phosphorylation of Rnc1 *in vivo*. The indicated cells were transformed with pREP1-GST-Rnc1 or pREP1-GST-Rnc1(T50A), and analysed for threonine phosphorylation. **c**, Pmk1 phosphorylates Rnc1 *in vitro*. GST-Rnc1 phosphorylation activity of the bacterially expressed and purified GST-Pmk1 was measured as described in Methods. **d**, Threonine (T50) phosphorylation is essential for the suppression of the Cl^- sensitivity of calcineurin deletion. $\Delta ppb1$ cells overexpressing

wild-type Rnc1 or its phosphorylation-site mutants were analysed as in Fig. 1a.

e, Pmk1-mediated phosphorylation regulates the RNA-binding activity of Rnc1. The indicated cells were analysed by immunoprecipitation and RT-PCR. **f**, Pmk1 signalling-responsive stabilization of Pmp1 mRNA. GST-Pmp1 mRNA stability was analysed in the indicated cells. **g**, Model for the mechanism of the RNA-binding protein-mediated feedback regulation of MAPK signalling. The activated Pmk1 phosphorylates Rnc1, leading to enhanced activity of Rnc1, thereby binding and stabilizing Pmp1 mRNA. Upregulated Pmp1 dephosphorylates and inactivates Pmk1, thus making a negative-feedback loop that regulates MAPK signalling.

Next, we examined whether Rnc1 is associated with Pmp1 mRNA *in vivo* at the endogenous level. In wild-type cell extracts, Rnc1 was efficiently immunoprecipitated using an anti-Rnc1 antibody, and we detected the endogenous Pmp1 mRNA by polymerase chain reaction with reverse transcriptase (RT-PCR) analysis of the immunoprecipitates (Fig. 3c). In contrast, Pmp1 mRNA was not detected in $\Delta pmp1$ nor $\Delta rnc1$ mutants (Fig. 3c). Thus, Rnc1 associates with Pmp1 mRNA *in vivo*.

We noted the presence of three repeats of the tetranucleotide UCAU, which is located 170 nucleotides downstream of the stop codon in the 3' untranslated region (3' UTR) of Pmp1 mRNA (Supplementary Fig. S1a). Notably, the Nova-1 KH-type RNA-binding protein binds to an element containing repeats of the tetranucleotide UCAU²⁰. We thus generated a series of Pmp1 mRNA mutants with mutated 3' UTRs (Supplementary Fig. S1a). The mRNA levels of the M1, M3 and M13 mutants were stable and indistinguishable from that of the wild type, whereas M2, M12, M23 and M123 mutants showed significantly reduced stability, with the triple mutant M123 being the least stable (Supplementary Fig. S1b). Consistently, M123 failed to suppress the Cl^- sensitivity of $\Delta ppb1$, and the ability of these 3' UTR mutants to suppress $\Delta ppb1$ correlated well with the stability of their mRNAs (Supplementary Fig. S1c). In addition, the overproduction of Rnc1 stabilized Pmp1 mRNA levels with the wild-type 3' UTR, whereas the M123 mutant exhibited minimal response to Rnc1 overexpression (Supplementary Fig. S1b). The M123 mutation abolished the binding of Rnc1 to Pmp1 mRNA (Fig. 3a, right panel). Moreover, overexpression of $rnc1^+$ failed to suppress the Cl^- sensitivity of $\Delta ppb1$ cells when the 3' UTR of $pmp1^+$ containing the UCAU triple repeats was deleted (Supplementary Fig. S1d). Thus, the UCAU repeats in the 3' UTR of Pmp1 mRNA are required for Rnc1 binding and for the regulation of Pmp1 mRNA stability by Rnc1.

Rnc1 open reading frame has six potential proline-directed MAPK phosphorylation sites (Ser-Pro/Thr-Pro (TP))²¹ (Fig. 4a). To examine their phosphorylation, we used an anti-phosphothreonine-proline (phospho-TP)-specific antibody. TP phosphorylation of GST-Rnc1 was detected in wild-type cells but not in $\Delta pmk1$ cells (Fig. 4b). Overexpression of Pek1^{DD} (Pek1(S234D/T238D)), the constitutively active form of MAPKK for Pmk1 (ref. 13), markedly increased the amount of TP phosphorylation on GST-Rnc1; overexpression of Pek1^A abolished the phosphorylation (Fig. 4b). Furthermore, the bacterially expressed and purified GST-Pmk1, but not the kinase-dead version of Pmk1 (Pmk1(K52R)), phosphorylated purified GST-Rnc1 *in vitro* (Fig. 4c). These results indicate that the observed phosphorylation of Rnc1 is, in fact, Pmk1-dependent and is not due to an adventitiously associated protein kinase.

To assess the functional significance of the phosphorylation sites in Rnc1, we constructed a series of substitution mutants by replacing either the Thr or Ser residue of these putative phospho-acceptor sites (Fig. 4a) with Ala, and expressed them in $\Delta ppb1$ cells. The expression of the Rnc1(T50A) mutant failed to suppress the Cl^- sensitivity of $\Delta ppb1$, whereas wild-type $rnc1^+$ and other mutant versions fully suppressed the sensitivity (Fig. 4d). Consistently, we found that the Rnc1(T50A) mutant was no longer phosphorylated (Fig. 4b). We also constructed Rnc1(T50D) and Rnc1(T50E) mutants by replacing Thr 50 with Asp and Glu, respectively. These phosphorylation-mimic mutants more potently suppressed the Cl^- sensitivity compared with wild-type Rnc1 (Fig. 4d). Thus, phosphorylation of Thr 50 is essential for the activity of Rnc1 in the stabilization of Pmp1 mRNA.

We also examined whether phosphorylation of Rnc1 by Pmk1 is able to regulate the ability of Rnc1 to bind and stabilize Pmp1 mRNA. The activation of Pmk1 by the overproduction of Pek1^{DD} markedly increased the amount of Pmp1 mRNA bound to Rnc1 (Fig. 4e). In contrast, inactivation of Pmk1 signalling by either the expression of Pek1^A or the deletion of the $pmk1^+$ gene abolished the

ability of Rnc1 to bind Pmp1 mRNA (Fig. 4e). In addition, overproduction of Pek1^{DD} stabilized Pmp1 mRNA levels, whereas the overproduction of Pek1^A or Pmk1 deletion destabilized Pmp1 mRNA levels (Fig. 4f). These observations are consistent with the idea that Rnc1 stabilizes Pmp1 mRNA in response to Pmk1 signalling.

We have shown that mRNA stabilization of a MAPK phosphatase is mediated by a new KH-type RNA-binding protein. Moreover, MAPK signalling-mediated phosphorylation controls the ability of the RNA-binding protein to stabilize the inherently unstable MAPK phosphatase mRNA, consequently leading to a negative-feedback loop of the MAPK signalling pathway (Fig. 4g). Some of the members of the MAPK phosphatase family, including CL100/MKP-1, are encoded by immediate-early genes, and exposure of the cell to growth factors, cytokines, cell stresses or activated oncogenes leads to the induction of a subset of MAPK phosphatase genes^{5,7,8}. This transcriptional upregulation of MAPK phosphatase is well documented as a mechanism of the negative-feedback regulation of MAPK signalling. Our study provides evidence for an additional negative-feedback mechanism of MAPK signalling mediated by an RNA-binding protein, thereby highlighting the complexity and importance of tight control of MAPK phosphatase. □

Methods

Gene disruption of $rnc1^+$

A 2-kilobase (kb) *PstI/NsiI* fragment containing $rnc1^+$ was subcloned into pGEM-5Zf to create pGEM-rnc1. pGEM-rnc1 was cleaved at the single *HindIII* site in $rnc1^+$ and ligated to *S. pombe ura4⁺* (1.8 kb). This construct was used to transform haploid cells²².

Northern and northwestern blot analyses

Total RNA was isolated by the method described in ref. 23. Ten micrograms of total RNA per lane was subjected to electrophoresis on denaturing formaldehyde 1% agarose gels and transferred onto nylon membranes. Hybridization was performed using DIG-labelled antisense cRNA probes for GST and Leu1. The DIG-labelled hybrids were detected by an enzyme-linked immunoassay using an anti-DIG-alkaline phosphatase antibody conjugate. The hybrids were visualized by chemiluminescence on a light-sensitive film in accordance with the manufacturer's instructions (Roche). Northwestern blot analysis was conducted as described²⁴. Pmp1 RNA probe for the northwestern blot analysis was synthesized from a Bluescript plasmid containing a 430 bp *HindIII/BamHI* fragment encoding the Pmp1 3' UTR. To analyse the Pmp1 mRNA stability, cells were transformed with a pREP41GST-Pmp1 3' UTR fusion gene, and cultured in EMM. The mRNA stability was determined from semi-logarithmic plots of the fraction of GST-Pmp1 mRNA remaining after the addition of thiamine versus time after thiamine addition. Fractions of mRNA remaining were normalized to the level of Leu1 mRNA to control for loading error. Similarly, the transcript levels of Pmp1 with mutant 3' UTR and GST-Pek1 (ref. 13) were determined. The semi-logarithmic graphs show the averages of three different experiments.

Immunoprecipitation and RT-PCR

Exponentially growing cells (2×10^8) were disrupted with glass beads in 200 μ l extraction buffer (50 mM Tris-HCl, pH 7.5, 150 mM KCl, 2 mM MgCl₂ containing 20 mM vanadyl ribonucleoside complexes, 200 U ml⁻¹ RNasin, 0.1% NP-40, 1 mM dithiothreitol, 0.2 mg ml⁻¹ heparin, and a mixture of protease inhibitors (1 mM phenylmethylsulphonyl fluoride, 0.1 mM benzamide, 0.1 mM sodium metabisulphite, 0.1 μ g ml⁻¹ chymostatin, 2 μ g ml⁻¹ aprotinin, 1 μ g ml⁻¹ pepstatin A, 1 μ g ml⁻¹ phosphoramidon and 0.5 μ g ml⁻¹ leupeptin)). Extracts were cleared by centrifugation (15 min at 15,000 g). An anti-Rnc1 antibody was added to the cleared extracts and incubated for 1 h on ice after incubation with protein A agarose for 1 h at 4 °C. Beads were washed four times in wash buffer (50 mM Tris-HCl, pH 7.5, 150 mM KCl, 2 mM MgCl₂) and were heated in 50 mM Tris-HCl, pH 8.0, 100 mM NaCl, 10 mM EDTA, 1% SDS for 10 min at 65 °C. Eluted samples were extracted with phenol-chloroform, ethanol precipitated, resuspended in DNase buffer and treated with RNase-free DNase. The remaining RNA was extracted, precipitated and resuspended in 20 μ l diethylpyrocabonate water. RT-PCR was performed with 1 μ l RNA as a template using Pmp1-specific primers. The number of amplification cycles was adjusted to avoid reaching a plateau phase during PCR.

Protein expression

For protein expression in yeast, thiamine-repressible *nmt1* promoter was used¹⁸. Expression was repressed by the addition of 4 μ g ml⁻¹ thiamine to EMM while expression was induced by washing and incubating the cells in EMM lacking thiamine. The GST-fused gene was subcloned into pREP1, or pREP41 vectors, obtaining a maximum expression of the fused gene using pREP1, whereas pREP41 contained the attenuated version of the *nmt1* promoter¹⁸. For the expression of Rnc1 as a GST-fusion protein, a *BamHI* fragment containing the coding sequence of $rnc1^+$ was amplified by PCR and inserted into the *BamHI* site of pREP1-GST⁹, to create pREP1-GST-Rnc1. For the expression of Pmp1 as a

GST-fusion protein, a *Bam*HI fragment containing the coding sequence and the 3' UTR was amplified by PCR and inserted into the *Bam*HI site of pREP41-GST, to create pREP41-GST-Pmp1. GST-fused proteins expressed in yeast cells were purified with glutathione beads before performing immunoblot or northwestern analyses.

For protein expression in bacteria, the full-length Pmk1 complementary DNA was amplified by RT-PCR from the wild-type fission yeast total RNA. GST-fusion proteins encoding Pmk1 and its kinase-dead version (Pmk1(K52R)) were constructed using pGEX-4T (Pharmacia), expressed in *Escherichia coli* DH5, and purified using glutathione-Sepharose beads as previously described². The purified GST-fusion proteins were used for *in vitro* kinase reactions as described below.

Site-directed mutagenesis and phosphorylation assay

Site-directed mutagenesis was performed using the QuickChange Site-Directed Mutagenesis Kit (Stratagene). Phosphorylation of Pmk1 was analysed using anti-phospho Pmk1 antibodies as described previously¹³.

In vitro kinase reactions were performed as previously described¹³ with some modifications. Bacterially expressed and purified GST-Pmk1 or GST-Pmk1(K52R) in place of Pmk1-haemagglutinin was used as an enzyme, and GST-Rnc1 purified from fission yeast in place of the myelin basic protein was used as a substrate.

Threonine phosphorylation of Rnc1 was analysed by immunoblotting using an antibody directed against phosphothreonine-proline (Cell Signalling Technology).

Received 10 February; accepted 11 July 2003; doi:10.1038/nature01907.

1. Herskowitz, I. MAP kinase pathways in yeast: for mating and more. *Cell* **80**, 187–197 (1995).
2. Marshall, C. J. MAP kinase kinase kinase, MAP kinase kinase and MAP kinase. *Curr. Opin. Genet. Dev.* **4**, 82–89 (1994).
3. Nishida, E. & Gotoh, Y. The MAP kinase cascade is essential for diverse signal transduction pathways. *Trends Biochem. Sci.* **18**, 128–131 (1993).
4. Levin, D. E. & Errede, B. The proliferation of MAP kinase signaling pathways in yeast. *Curr. Opin. Cell Biol.* **7**, 197–202 (1995).
5. Camps, M., Nichols, A. & Arkinstall, S. Dual specificity phosphatases: a gene family for control of MAP kinase function. *FASEB J.* **14**, 6–16 (2000).
6. Keyse, S. M. Protein phosphatases and the regulation of mitogen-activated protein kinase signalling. *Curr. Opin. Cell Biol.* **12**, 186–192 (2000).
7. Sun, H., Charles, C. H., Lau, L. F. & Tonks, N. K. MKP-1 (3CH134), an immediate early gene product, is a dual specificity phosphatase that dephosphorylates MAP kinase *in vivo*. *Cell* **75**, 487–493 (1993).
8. Keyse, S. M. & Emslie, E. A. Oxidative stress and heat shock induce a human gene encoding a protein-tyrosine phosphatase. *Nature* **359**, 644–647 (1992).
9. Sugiura, R., Toda, T., Shuntoh, H., Yanagida, M. & Kuno, T. *pmp1*⁺, a suppressor of calcineurin deficiency, encodes a novel MAP kinase phosphatase in fission yeast. *EMBO J.* **17**, 140–148 (1998).
10. Toda, T. *et al.* The fission yeast *pmp1*⁺ gene encodes a novel mitogen-activated protein kinase homolog which regulates cell integrity and functions coordinately with the protein kinase C pathway. *Mol. Cell Biol.* **16**, 6752–6764 (1996).
11. Zaitsevska-Carter, T. & Cooper, J. A. Spm1, a stress-activated MAP kinase that regulates morphogenesis in *S. pombe*. *EMBO J.* **16**, 1318–1331 (1997).
12. Yoshida, T., Toda, T. & Yanagida, M. A calcineurin-like gene *ppb1*⁺ in fission yeast: mutant defects in cytokinesis, cell polarity, mating and spindle pole body positioning. *J. Cell Sci.* **107**, 1725–1735 (1994).
13. Sugiura, R., Toda, T., Dhut, S., Shuntoh, H. & Kuno, T. The MAPK kinase Pek1 acts as a phosphorylation-dependent molecular switch. *Nature* **399**, 479–483 (1999).
14. Siomi, H., Matunis, M. J., Michael, W. M. & Dreyfuss, G. The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. *Nucleic Acids Res.* **21**, 1193–1198 (1993).
15. Burd, C. G. & Dreyfuss, G. Conserved structures and diversity of functions of RNA-binding proteins. *Science* **265**, 615–621 (1994).
16. Ostareck-Lederer, A., Ostareck, D. H. & Hentze, M. W. Cytoplasmic regulatory functions of the KH-domain proteins hnRNPs K and E1/E2. *Trends Biochem. Sci.* **23**, 409–411 (1998).
17. Loewith, R., Hubberstey, A. & Young, D. Skh1, the MEK component of the Mkh1 signaling pathway in *Schizosaccharomyces pombe*. *J. Cell Sci.* **113**, 153–160 (2000).
18. Maundrell, K. Thiamine-repressible expression vectors pREP and pRIP for fission yeast. *Gene* **123**, 127–130 (1993).
19. Lewis, H. A. *et al.* Sequence-specific RNA binding by a Nova KH domain: implications for paraneoplastic disease and the fragile X syndrome. *Cell* **100**, 323–332 (2000).
20. Buckanovich, R. J. & Darnell, R. B. The neuronal RNA binding protein Nova-1 recognizes specific RNA targets *in vitro* and *in vivo*. *Mol. Cell Biol.* **17**, 3194–3201 (1997).
21. Davis, R. J. The mitogen-activated protein kinase signal transduction pathway. *J. Biol. Chem.* **268**, 14553–14556 (1993).
22. Rothstein, R. J. One-step gene disruption in yeast. *Methods Enzymol.* **101**, 202–211 (1983).
23. Kohrer, K. & Domdey, H. Splicing and spliceosome formation of the yeast MATa1 transcript require a minimum distance from the 5' splice site to the internal branch acceptor site. *Nucleic Acids Res.* **6**, 9457–9475 (1988).
24. Rodriguez, P. L. & Carrasco, L. Nonradioactive northwestern analysis using biotinylated riboprobes. *Biotechniques* **17**, 702–707 (1994).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank M. Yanagida and T. Toda for their gift of strains and plasmids, S. Nakielnny and M. Sakaue for discussions, and M. Takeuchi for technical assistance. This work was supported in part by research grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.S. (sugiur@med.kobe-u.ac.jp).

Transcription factor IIB acetylates itself to regulate transcription

Chu H. Choi, Makoto Hiromura & Anny Usheva

Endocrinology Division, Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, 99 Brookline Avenue, Boston, Massachusetts 02215, USA

Acetylation is a well-known regulatory post-translational modification¹, but a biological function for acetylation in regulating basal transcription factors has not been reported. Here we show that the general transcription factor TFIIB, which is required for the initiation of eukaryotic polymerase II transcription², is acetylated. TFIIB is also an autoacetyltransferase, although it shares no sequence homology with any known acetyltransferases. In the absence of other enzymes, it binds acetyl-coenzyme A (acetyl-CoA), and catalyses the transfer of the acetyl group onto a specific lysine residue (K238). Both recombinant and cellular TFIIB can autoacetylate, markedly stabilizing the interaction between TFIIB and transcription factor TFIIF and activating transcription *in vitro* and in cells. A K238A mutant, which cannot be autoacetylated, does not show this activation of transcription. Our findings suggest that there is a regulatory pathway controlling acetylation of TFIIB, and they link acetyl-CoA with basal gene transcription.

Eukaryotic self-catalytic acyl transfer activity has been reported in a few enzymes involved in lipid metabolism^{3–6}. However, autozyme activity has not been reported for any eukaryotic transcription factor. Here we show that TFIIB functions as an enzyme, with itself serving as the substrate, and can be considered an autozyme. Incubation of highly purified recombinant human TFIIB (rhTFIIB) with labelled [¹⁴C]-acetyl-CoA at ambient temperature results in the formation of an acetyl-TFIIB derivative (Fig. 1). After incubation, the protein was denatured, isolated and tested for the incorporation of radioactivity. The addition of acetyl groups to TFIIB depends on the concentration of both [¹⁴C]-acetyl-CoA (Fig. 1a) and TFIIB (Fig. 1b), suggesting their direct interaction through this process. Because it is likely that the structure of TFIIB is important to its interaction with acetyl-CoA, denaturing the protein ensures that the result reflects not merely binding of [¹⁴C]-acetyl-CoA, but a covalent attachment of radioactive acetyl groups to the protein. The purity of the TFIIB protein (see Supplementary Information) makes it very unlikely that any contaminating prokaryotic proteins with acetyltransferase activity caused this autoacetylation of TFIIB.

In order to determine the specificity of the autoacetylation reaction, other basic transcription factors were tested for self-acetylation. The experiment was repeated using recombinant human TATA box-binding protein (TBP), yeast TFIIB, human Yin Yang 1 (YY1), bovine serum albumin (BSA) and rhTFIIB. Acetylation occurs only with yeast and human TFIIB (Fig. 1c). The conservation of this autoacetylation activity in both yeast and humans may suggest an important physiological role for the acetylation of TFIIB.

We tested if TFIIB could act as a histone acetyltransferase (HAT). Incubation of TFIIB with [¹⁴C]-acetyl-CoA and histone octamer assembled from highly purified, bacterially produced histones did not result in histone acetylation (Fig. 1d). Autoradiography clearly shows that recombinant TFIIB is acetylated, whereas none of the histone proteins show incorporation of radioactivity. As histones are known to be acetylated by HATs⁷, this strengthens the claim that this reaction is indeed autoacetylation, and not the result of contamination by an acetyltransferase.

Fractionated nuclear extract from HeLa cells was analysed in an

attempt to isolate acetylated TFIIB (see Supplementary Information). Immunoprecipitation with antibodies specific for acetylated lysine reveals that a subset of the nuclear TFIIB is indeed acetylated in cells. It appears that in nuclear extracts from HeLa (see Supplementary Information) and human primary smooth muscle cells (not shown), approximately 10–15% of TFIIB is in acetylated form. The specificity of the α -acetyllysine antibody for acetylated rhTFIIB was verified by immunoaffinity chromatography and western blotting (see Supplementary Information).

TFIIB autoacetylation activity was also verified using an immobilized on-membrane acetylation procedure (Fig. 1e). We observed [14 C]-labelled acetylated recombinant TFIIB and a 34-kilodalton (34-kDa) acetylated product in the nuclear fraction, which is the size of wild-type TFIIB. Western blotting verified the presence of TFIIB in all three experiments at the same position as the [14 C] labelled products. In the case of cellular extract pre-treated with non-radioactive acetyl-CoA before SDS–polyacrylamide gel electrophoresis (SDS–PAGE), no further acetylation is seen, as the

TFIIB is probably already in acetylated form (Fig. 1e, lane 6). The [14 C]-acetyl TFIIB product formed at ambient temperature, coming to completion after 30 min (Fig. 2a). We observed that when TFIIB was first denatured, no acetylation activity was detected (Fig. 2a, lane 1). This demonstrates that the structure-based association of TFIIB and acetyl-CoA is a necessary step to this autoacetylation process.

Acetyl-CoA binds strongly to TFIIB in solution. Assuming that the binding is the limiting factor for reaction rate, the interaction of TFIIB and acetyl-CoA fits the Michaelis–Menten model of enzyme kinetics. A linear Lineweaver–Burk plot demonstrates a Michaelis constant, K_m , of 0.117 mM for acetyl-CoA and an inhibition constant, K_i , of 0.417 mM for CoA (Fig. 2b). The best linear fit was obtained with concentrations above 0.5 mM of acetyl-CoA. The binding of acetyl-CoA to TFIIB became evident in our experiments in other ways as well. After extensive dialysis of a solution of TFIIB and acetyl-CoA, there was still a significant amount of acetyl-CoA, and the two moieties even eluted through high-performance liquid chromatography (HPLC) together (not shown). The coincident values of V_{max} (reaction at infinite concentration of substrate) for acetyl-CoA and coenzyme A (Fig. 2b) suggest a specific, common binding site for the two similar cofactors on the surface of TFIIB. Thus, it is likely that CoA functions as a competitive inhibitor of acetyl-CoA. This was verified by demonstrating that the presence of CoA inhibits the autoacetylation reaction significantly (see Supplementary Information).

These data support the hypothesis that TFIIB is an enzyme, with acetyl-CoA as the requisite cofactor. The reaction between rhTFIIB and acetyl-CoA results in the generation of unacetylated CoA (Fig. 2c, lane 4). After 40 min, the reaction is complete and no detectable acetyl-CoA is left. In the presence of denatured TFIIB, less than 5% degradation of acetyl-CoA was detected (Fig. 2c, lane 5).

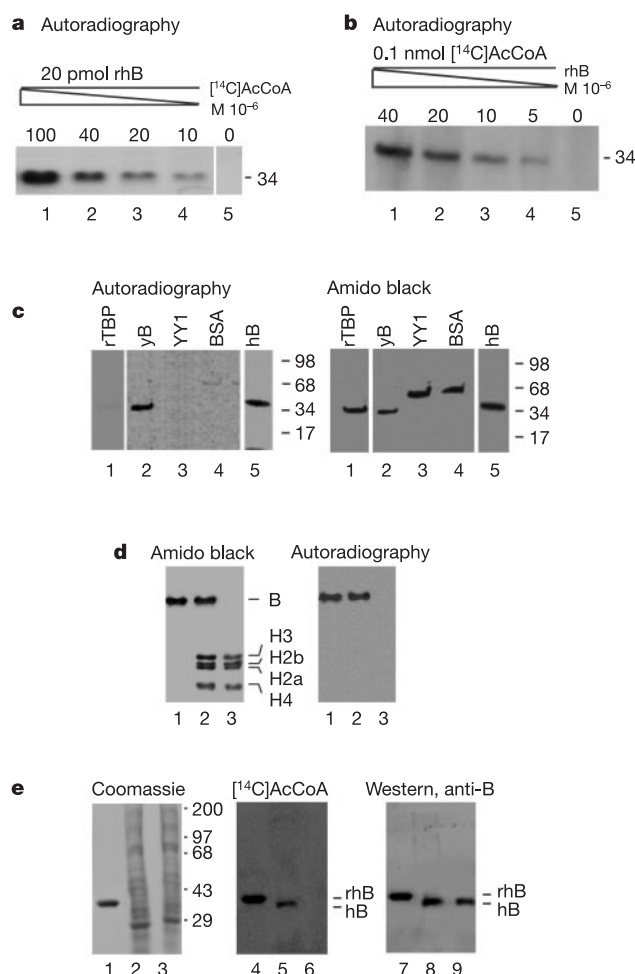


Figure 1 TFIIB is an autoacetyltransferase. **a**, **b**, Autoacetylation of recombinant human TFIIB (rhB) depends on the concentration of both acetyl-CoA (**a**) and TFIIB (**b**). **c**, Autoacetylation conditions applied to highly purified recombinant proteins: lane 1, human TBP (rTBP); lane 2, yeast TFIIB (yB); lane 3, human YY1; lane 4, BSA (Gibco); lane 5, human TFIIB (hB). **d**, Autoacetylation conditions applied to rhTFIIB and recombinant nucleosome histone octamer. Lane 1, rhTFIIB; lane 2, rhTFIIB and 10 pmol histone octamer; lane 3, 10 pmol histone octamer. **e**, Acetylation of proteins from HeLa nuclear extract fraction. Lanes 1, 4, 7, (His)₈-hTFIIB; lanes 2, 5, 8 nuclear extract fraction; lanes 3, 6, 9 nuclear extract fraction pretreated with 5 mM of unlabelled acetyl-CoA for 40 min before electrophoresis. Anti-B, anti-TFIIB.

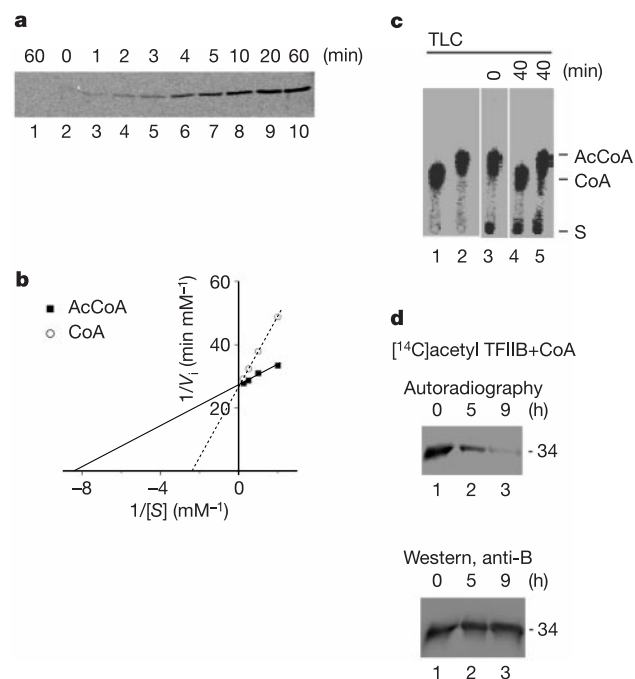


Figure 2 Kinetic analysis of TFIIB autoacetylation. **a**, The extent of acetylation as a function of time at fixed TFIIB and [14 C]-acetyl-CoA concentrations. Lane 1, SDS-denatured TFIIB incubated with acetyl-CoA. **b**, Lineweaver–Burk plot of TFIIB binding with acetyl-CoA and CoA. $[S]$ represents the concentration of acetyl-CoA or CoA. **c**, Acetylation reaction mixture separation by TLC. Lane 1, 50 nmol of CoA; lane 2, 50 nmol of AcCoA; lane 3, acetylation reaction at 0 min; lane 4, reaction after 40 min; lane 5, heat and SDS-denatured hTFIIB. **d**, Fully acetylated TFIIB is deacetylated by excess CoA over several hours.

These data suggest that TFIIB binds acetyl-CoA, transfers the acetyl group to itself, and then releases deacetylated CoA.

The autoacetylation of TFIIB is a reversible process. Fully acetylated TFIIB was placed in an excess of CoA and left for several hours. After 5 h, a large decrease in TFIIB radioactivity was recorded, and after 9 h, the signal was reduced by 85% (Fig. 2d). The acetylation process is reversible in the presence of CoA, although deacetylation of TFIIB proceeds at a much slower rate than acetylation. Presumably, CoA binds to acetyl-TFIIB, and the acetyl group can be transferred back to CoA, though it does not appear to be thermodynamically favoured. This equilibrium can be overcome by a large excess of CoA, as our experiment demonstrates. This reversibility shows that the acetylation is a truly catalytic process. Regulatory modifications should be reversible to be highly flexible, as in the case of phosphorylation and dephosphorylation. The observation of deacetylation of TFIIB lends credence to the idea that this auto-

acetylation may act as a regulatory signal in the function of TFIIB *in vivo*.

The number and location of acetylation sites in the autoacetylation enzymatic reaction were determined. Dialysed acetylated rhTFIIB was denatured and digested with trypsin to give peptide fragments, and analysed by liquid chromatography-mass spectrometry/mass spectrometry (LC-MS/MS)^{8–10}. Only one lysine was unreactive to cleavage because of N-acetylation, at amino-acid position 238 (Fig. 3a). NMR¹¹ and crystal structures¹² of TFIIB show K238 to be on the surface of the protein, exposed to the solvent and therefore accessible to acetylation. It is notable that lysine is preferentially acetylated over cysteine, which is the target of previously reported autoacetylation¹³.

In order to verify that K238 is indeed the only acetylated residue in the protein, a K238A human TFIIB mutant (K238A) was produced and subjected to the acetylation conditions. The incorporation of radioactivity vanishes upon the substitution of alanine for this lysine residue (Fig. 3b, lane 2). As a control, we produced a K226A mutant in a nearby lysine residue situated on the same α -helix as K238, and also exposed to the solvent (Fig. 3c). This mutation did not affect the acetylation activity (Fig. 3b, lane 3). The reactive K238 and unreactive K226 are located on a surface of TFIIB that is free of interaction with TBP or DNA (Fig. 3c)¹⁴. Therefore, acetylation of TFIIB would not be expected to influence interactions with these partners. However, this surface might be involved in other protein interactions.

To identify altered protein–protein interactions of acetyl-TFIIB, experiments were performed using glutathione *S*-transferase (GST)–hTFIIB and acetylated GST–hTFIIB affinity capturing of other highly purified recombinant transcription factors. Western blots demonstrated that acetylation of TFIIB does not influence its interaction with TBP, as predicted¹² (Fig. 4a). However, acetyl-TFIIB beads captured 90% of TFIIF, whereas the unacetylated form recruited only 15% of the available TFIIF. This change in the TFIIB–TFIIF interaction is great enough to affect transcription initiation significantly. To assess the effects of acetyl-TFIIB on transcription, *in vitro* transcription reactions were performed using TFIIB-depleted nuclear extracts from HeLa cells, supplementing with rhTFIIB, acetyl-rhTFIIB, K238A mutant TFIIB (K238A), and acetyl-K238A. Compared to wild-type TFIIB (Fig. 4b, lanes 3–5), acetylated TFIIB strongly activates transcription (Fig. 4b, lanes 6–8). K238A, whether subjected to acetylating conditions or not, shows roughly the same behaviour as wild-type TFIIB (Fig. 4b, lanes 3–8). This is due to the fact that alanine at the 238 position in the mutant TFIIB cannot be acetylated.

To verify these effects on transcription in cells, we devised transfection experiments using reporter vectors coupled to the promoters AdMLP (adenovirus major late promoter), RSV (Rous sarcoma virus) and SV40 (simian virus 40) with different amounts of DNA constructs for the overexpression of either wild-type flag-tagged TFIIB or flag-tagged K238A (Fig. 4c). Referencing relative reporter levels, we see that in all cases, overexpressing K238A TFIIB mutant represses transcription, whereas rhTFIIB boosts it (Fig. 4c). The repression caused by the mutant TFIIB, which cannot be acetylated, may be understood from the standpoint of competition with endogenous TFIIB, which can be acetylated, for transcription initiation complexes. Western blot analysis verified the presence of ectopically expressed proteins (Fig. 4d). We found that a portion of the overexpressed flag-TFIIB is acetylated, whereas flag-tagged K238A is not acetylated in cells (see Supplementary Information). Chromatin immunoprecipitation (ChIP) assays verified the presence of flag-TFIIB and K238A flag-TFIIB at the AdMLP, suggesting that the overexpressed proteins do participate in the transcription of the promoter (Fig. 4e). When ChIP analysis is applied to cells overexpressing K238A flag-TFIIB, less than 30% of endogenous TFIIF is captured on the AdMLP compared to cells overexpressing wild-type flag-TFIIB (Fig. 4e).

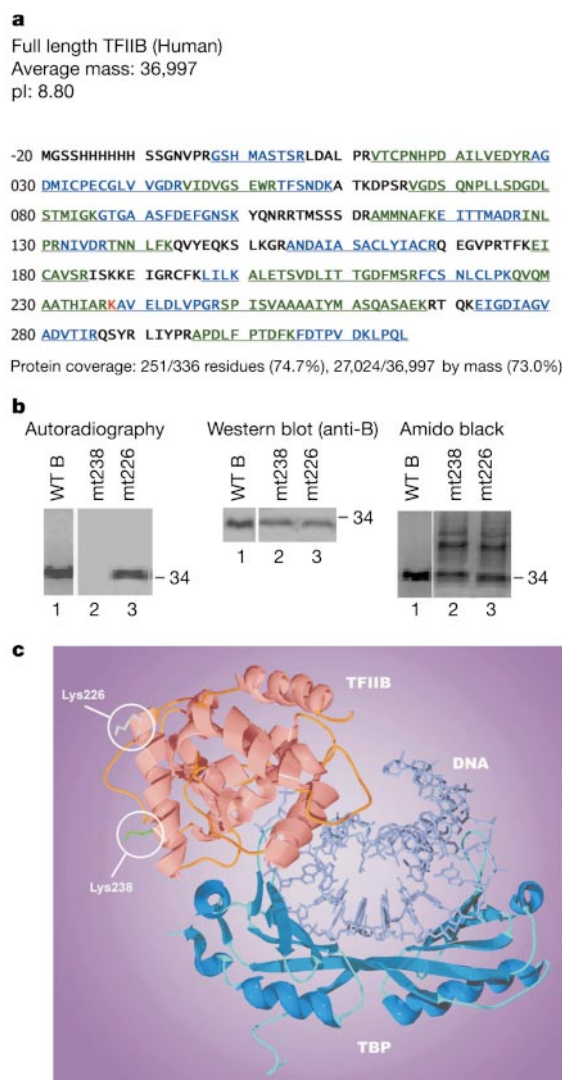


Figure 3 The acetylation site is localized on K238 of the hTFIIB core domain. **a**, Using LC-MS/MS, the lone acetylated lysine 238 is presented in red (K). **b**, Autoacetylation with wild-type (WT) and mutant (m) hTFIIB variants. The presence of TFIIB and its variants was verified by western blot (anti-B) and a secondary antibody-peroxidase derived coloured reaction. Lane 1, highly purified wild-type rhTFIIB (WT B); lane 2, partially purified K238A mutant (mt238); lane 3, partially purified K226A mutant (mt226). **c**, Ribbon diagram of the ternary complex crystal structure of TFIIB (red), TBP (blue) and a TATA DNA sequence¹³. The reactive K238 (green) and unreactive K226 (grey) are highlighted. The diagram was made using Swiss PDB Viewer, POV-Ray 3.1 and Adobe Photoshop.

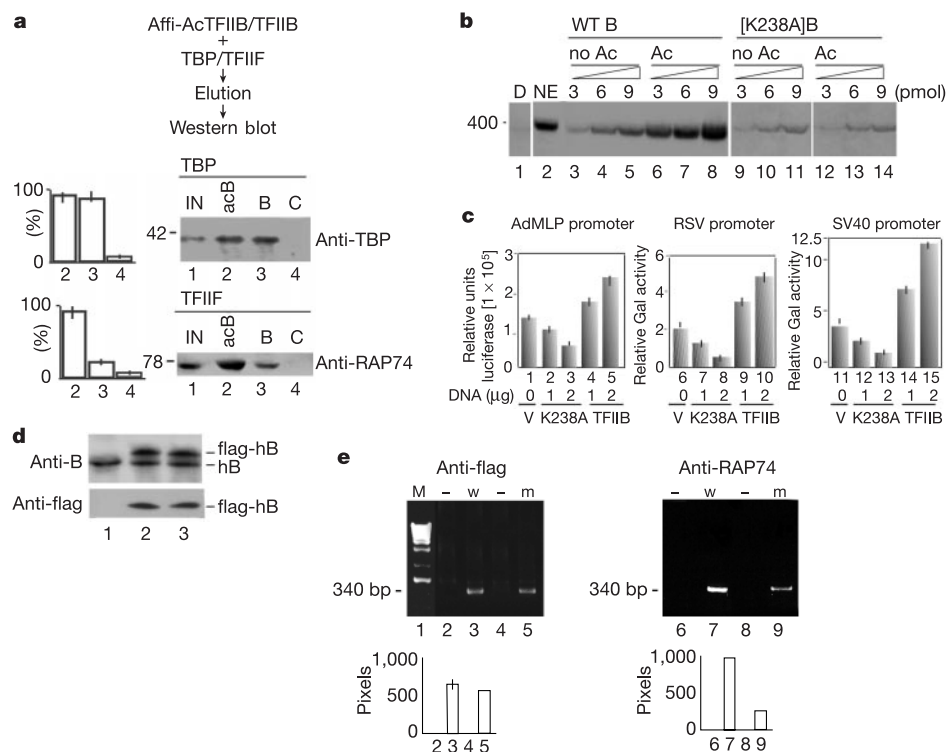


Figure 4 Function of acetylated hTFIIB. **a**, Immunoaffinity capture of TBP and TFIIF by acetylated and unacetylated rhTFIIB. Lane 1, input (IN) at one-third concentration; lane 2, acetylated rhTFIIB affinity matrix (acB); lane 3, unacetylated rhTFIIB affinity matrix (B); lane 4, control affinity matrix without TFIIB. **b**, Transcription assays *in vitro*. Lane 1, depleted nuclear extract (D); lane 2, undepleted nuclear extract (NE). Lanes 3–8, depleted nuclear extract supplemented with rhTFIIB (no Ac) or acetyl-rhTFIIB (Ac). Lanes 9–14, depleted nuclear extract supplemented with K238A without (no Ac) and with acetylation conditions (Ac). **c**, Transient transfection. Ectopic expression of mutant K238A TFIIB dose-dependently inhibits transcription. Lanes 1, 6, 11, 0.5 μ g reporter plasmid DNA (V); lanes 2, 7, 12, reporter and 1 μ g K238A plasmid; lanes 3, 8, 13, reporter and 2 μ g K238A plasmid; lanes 4, 9, 14, reporter and 1 μ g rhTFIIB plasmid; lanes 5, 10, 15, reporter and

2 μ g rhTFIIB plasmid. **d**, Wild-type and mutant TFIIB overexpression in transfected MBP cells. Lane 1, reporter vector only; lane 2, reporter vector and 2 μ g TFIIB vector; lane 3, reporter vector and 2 μ g K238A vector. **e**, Chromatin immunoprecipitation (ChIP) assays were used to determine the presence of both recombinant wild-type (w) and mutant (m) TFIIB and endogenous TFIIF at the AdMLP promoter. Lane 1, DNA marker (M); lanes 2 and 3, wild type TFIIB co-transfected cell lysate with unrelated antibody (–) and anti-flag (w); lanes 4 and 5, K238A mutant TFIIB co-transfected cell lysate with unrelated antibody (–) and anti-flag (m); lanes 6 and 7, wild-type TFIIB co-transfected cell lysates with unrelated antibody (–) and anti-RAP74 (w); lanes 8 and 9, K238A mutant TFIIB co-transfected cell lysate with unrelated antibody (–) and anti-RAP74.

It is known that the TFIIB–TFIIF interaction participates in the formation of the transcriptional preinitiation complex¹⁵. It is likely that the increased stability of this interaction, as a consequence of TFIIB acetylation, contributes to the activated transcription. How TFIIB acetylation affects other higher-order interactions, for example, with RNA polymerase II, is of great importance, and remains to be determined. We note that it has been widely held that acetyltransferases are required to perform protein acetylation modifications on transcription factors¹. However, TFIIB is auto-acetylated in the presence of acetyl-CoA without the assistance of an acetyltransferase or related enzyme. Also, acetyl-TFIIB can be deacetylated without the aid of any deacetylases. As autoacetylation is a newly identified autoenzymatic transformation in transcription factors, it is not yet known whether or not other proteins that function in gene transcription are capable of similar self-modification. Such changes in cells hold unknown consequences for transcription regulation.

Finally, attachment of an acetyl group to the terminal amine of lysine results in the neutralization of one positive charge, and a shift in the local hydrophobicity of the protein surface. It is noteworthy that such a small post-translational modification on one component of the transcriptional complex could result in a significant transcriptional effect. The acetylation of TFIIB may be a new regulatory pathway in transcription. As acetyl-CoA is an important cofactor in this reaction, this suggests that basic transcription may

be linked to fatty-acid synthesis and cellular respiration. The implications of such connections would be far-reaching in cellular biology. □

Methods

Autoacetylation conditions

Highly purified rhTFIIB or nuclear extract fraction was incubated at ambient temperature for 30 min with [¹⁴C]-acetyl-CoA in the presence of 150 mM NaCl, 1 mM EDTA and 50 mM HEPES buffer at pH 7.4. The resulting products were denatured using 0.1% SDS and 1.0 μ l 2-mercaptoethanol, then separated by gel electrophoresis and transferred onto a nitrocellulose membrane. Protein bands were viewed by staining with Amido black or Coomassie blue. The membrane was stored overnight in a phosphor cassette (Molecular Dynamics) and the radioactive bands were visualized using a Molecular Dynamics 400-B Phosphorimager.

Nucleosome core particles were assembled¹⁶ from individual recombinant *Xenopus laevis* histones which were expressed in bacteria and chromatographically purified. The four histone expression vectors were a gift from T. Richmond.

Immobilized on-membrane acetylation

Recombinant TFIIB and a fraction of HeLa nuclear extract were subjected to SDS-PAGE and transferred to a nitrocellulose membrane for renaturing¹⁴. The membrane was soaked in PBS and 0.1% NP40 for 24 h, then exposed to acetylation conditions. Western blotting with anti-TFIIB antibodies ensures that TFIIB was acetylated.

Binding affinity assays

Binding of rhTFIIB to acetyl-CoA and CoA was measured by the reduction in absorbance at 260 nm in protein-filtered solution. All measurements were performed in triplicate, with s.d. < 0.006. K_m was taken as the negative reciprocal of the x-intercept of the linear fit to the data, and V_{max} is the reciprocal of the y-intercept.

Thin-layer chromatography

TLC was done on silica gel plates as previously described¹⁷. Acetyl-CoA and CoA were visualized by UV at 254 nm.

Deacetylation of acetyl-TFIIB

0.1 nmol of [¹⁴C]-acetyl-CoA was incubated with 20 pmol rhTFIIB for 40 min, allowing full acetylation. A 50-fold excess of CoA over initial acetyl-CoA was added and the mixture was incubated for the number of hours shown. The results were visualized by SDS-PAGE and autoradiography. The integrity of rhTFIIB after incubation was verified by western blotting with anti-TFIIB antibodies.

Proteolytic determination of acetylation sites

The acetylation site was determined by proteolytic digestion and LC-MS/MS analysis. Protein bands were excised from gels and subjected to limited tryptic digestion, followed by nanoscale reverse-phase HPLC separation of the resulting peptides. Peptides were identified after electrospray ionization into a LCQ DECA ion-trap mass spectrometer. For a peptide coverage of 74% of the protein, K238 is the only acetylated lysine residue.

Site-specific mutagenesis

The construct was mutated using the QuikChange kit (Stratagene) and custom primers (Oligos, Etc.). Mutant constructs were verified by sequencing. Proteins were expressed in *Escherichia coli* and purified as previously described¹². The presence of TFIIB and its variants was verified by western blot with anti-TFIIB antibodies and a secondary antibody- peroxidase derived coloured reaction.

Immunoaffinity capturing

TFIIB affinity matrices were prepared as previously described¹⁴. TFIIB and acetyl-TFIIB affinity agarose beads were mixed with highly purified recombinant TFIIB, then with rhTBP or TFIIF (Calbiochem) at ambient temperature. Binding to the TFIIB affinity matrix was monitored by western blotting with mouse monoclonal antibody against TBP (a gift from J. Flint) or rabbit polyclonal antibodies against RAP74 (C-18, Santa Cruz).

In vitro transcription assays

HeLa nuclear extract was depleted of TFIIB using an affinity matrix with covalently attached polyclonal antibody for hTFIIB. Transcription reactions were assembled with depleted nuclear extract, AdMLP G-less supercoiled DNA template¹⁸ and [³²P]-UTP. To substitute for endogenous TFIIB different amounts of highly purified rhTFIIB or K238A were added. Activity was then quantified by autoradiography after SDS-PAGE; the 340-nucleotide RNA transcript was generated from AdMLP.

Transfection experiments

Plasmids were purified using a CsCl gradient after Triton X14 treatment and transfected using Eugene 6 (Roche). Cell cultures were treated with lysis buffer 48 h post-transfection, and the luciferase reporter assay system (Promega Corp.) was used with the AdMLP. A β -galactosidase reporter assay system (Tropix) was used with the RSV and SV40 promoters. Cells received 3 μ g total plasmid DNA. Co-transfection of 0.5 μ g plasmid containing the luciferase gene in the pGL3-45S RNA promoter-EMCV vector (a gift from T. Moss), which does not require TFIIF for transcription, was used as a control to correct for varying transfection efficiencies. Galactosidase data are normalized to the activity seen with the plasmid promoter vector alone $\pm$ s.e.m. Transfection experiments were performed in four triplicate sets, with consistent results (s.d. $\pm$ 3%).

ChIP assays

The amount of TFIIF and flag-TFIIB variants attached to the co-transfected AdMLP in MBP cells was measured by ChIP assay, using a protocol from Upstate Solutions (with slight modifications). The chromatin was not sonicated. T7 and luciferase primers were used for the PCR amplification of the AdMLP fragment.

Received 12 May; accepted 1 July 2003; doi:10.1038/nature01899.

1. Kouzarides, T. Acetylation: A regulatory modification to rival phosphorylation? *EMBO J.* **19**, 1176–1179 (2000).
2. Orphanides, G., Lagrange, T. & Reinberg, D. The general transcription factors of RNA polymerase II. *Genes Dev.* **10**, 2657–2683 (1996).
3. Bizzozero, O. A., McGarry, J. F. & Lees, M. B. Autoacylation of myelin proteolipid protein with acyl coenzyme A. *J. Biol. Chem.* **262**, 13550–13557 (1987).
4. Yamashita, A., Watanabe, M., Tonegawa, T., Sugiura, T. & Waku, K. Acyl-CoA binding and acylation of UDP-glucuronosyltransferase isoforms of rat liver: Their effect on enzyme activity. *Biochem. J.* **312**, 301–308 (1995).
5. Duncan, J. A. & Gilman, A. G. Autoacylation of G protein alpha subunits. *J. Biol. Chem.* **271**, 23594–23600 (1996).
6. Veit, M. et al. Palmitoylation of rhodopsin with S-protein acyltransferase: Enzyme catalyzed reaction versus autocatalytic acylation. *Biochim. Biophys. Acta* **1394**, 90–98 (1998).
7. Brownell, J. E. & Allis, C. D. Special HATs for special occasions: Linking histone acetylation to chromatin assembly and gene activation. *Curr. Opin. Genet. Dev.* **6**, 176–184 (1996).
8. Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **68**, 850–858 (1996).
9. Peng, J. & Gygi, S. P. Proteomics: The move to mixtures. *J. Mass Spectrom.* **36**, 1083–1091 (2001).
10. Eng, J., McCormack, A. L. & Yates, J. R. I. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976–989 (1994).
11. Bagby, S. et al. Solution structure of the C-terminal core domain of human TFIIB: Similarity to cyclin A and interaction with TATA-binding protein. *Cell* **82**, 857–867 (1995).

12. Nikolov, D. B. et al. Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature* **377**, 119–128 (1995).
13. Yan, Y., Harper, S., Speicher, D. W. & Marmorstein, R. The catalytic mechanism of the ESA1 histone acetyltransferase involves a self-acetylated intermediate. *Nature Struct. Biol.* **9**, 862–869 (2002).
14. Usheva, A. & Shenk, T. YY1 transcriptional initiator: protein interactions and association with a DNA site containing unpaired strands. *Proc. Natl Acad. Sci. USA* **93**, 13571–13576 (1996).
15. Fang, S. M. & Burton, Z. F. RNA polymerase II-associated protein (RAP) 74 binds transcription factor (TF) IIB and blocks TFIIB-RAP30 binding. *J. Biol. Chem.* **271**, 11703–11709 (1996).
16. Luger, K., Rechsteiner, T. J. & Richmond, T. J. Preparation of nucleosome core particle from recombinant histones. *Methods Enzymol.* **304**, 3–19 (1999).
17. Mayer, R. T., Holman, G. M. & Bridges, A. C. Phosphorescence detection method for purine-containing compounds on thin-layer chromatograms. *J. Chromatogr.* **90**, 390–391 (1974).
18. Sawadogo, M. & Roeder, R. G. Factors involved in specific transcription by human RNA polymerase II: Analysis by a rapid and quantitative in vitro assay. *Proc. Natl Acad. Sci. USA* **82**, 4394–4398 (1985).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank T. Richmond for the gift of recombinant human histone proteins, S. Gygi for proteolytical analysis on acetyl-TFIIB, and S. Robson and M. Gray for critically reading our manuscript. This work was supported by the NIH (A.U.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.U. (ausheva@caregroup.harvard.edu).

Complex between nidogen and laminin fragments reveals a paradigmatic β -propeller interface

Junichi Takagi^{1,3*}, Yuting Yang^{2,4*}, Jin-huan Liu^{2,4}, Jia-huai Wang^{2,3} & Timothy A. Springer^{1,5}

¹The Center for Blood Research, Boston, Massachusetts 02115, USA

²Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA

³Department of Pediatrics, ⁴Department of Medicine, and ⁵Department of Pathology, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115, USA

* These authors contributed equally to this work

Basement membranes are fundamental to tissue organization and physiology in all metazoans. The interaction between laminin and nidogen is crucial to the assembly of basement membranes^{1–4}. The structure of the interacting domains reveals a six-bladed Tyr-Trp-Thr-Asp (YWTD) β -propeller domain in nidogen bound to laminin epidermal-growth-factor-like (LE) modules III3–5 in laminin (LE3–5). Laminin LE module 4 binds to an amphitheatre-shaped surface on the pseudo-6-fold axis of the β -propeller, and LE module 3 binds over its rim. A Phe residue that shutters the water-filled central aperture of the β -propeller, the rigidity of the amphitheatre, and high shape complementarity enable the construction of an evolutionarily conserved binding surface for LE4 of unprecedentedly high affinity for its small size⁵. Hypermorphic mutations in the Wnt co-receptor LRP5 (refs 6–9) suggest that a similar YWTD β -propeller interface is used to bind ligands that function in developmental pathways. A related interface, but shifted off-centre from the pseudo-6-fold axis and lacking the shutter over the central aperture, is used in the low-density lipoprotein receptor for an intramolecular interaction that is regulated by pH in receptor recycling¹⁰.

The minimal elements of basement membranes seem to be nidogen (entactin), laminin, type IV collagen and perlecan, because they are the only components universally conserved in

Caenorhabditis elegans, *Drosophila* and Chordata¹. The independent, self-assembling matrices formed by laminins and type IV collagens are linked together by nidogen, which also binds perlecan^{2,3}. Nidogen therefore seems to be central in organizing all basal laminae including those in skin, muscle and the nervous system. Laminin and nidogen are secreted by epithelial and mesenchymal cells, respectively², and their interaction is important in organogenesis²⁻⁴.

Laminin binding to nidogen has been localized to the third of three globular domains, G3, and its flanking epidermal growth factor (EGF)-6 domain (Fig. 1A)^{2,3}. The nidogen G3 domain contains six 45-residue sequence repeats with a YWTD motif that have been predicted to fold into a six-bladed β -propeller domain¹¹. The binding site within laminin has been localized to module LE4 in the γ 1 subunit^{2,12-14}.

Co-expression in 293T cells with a LE3-5 fragment of the laminin γ 1 subunit enhanced the secretion of His-tagged nidogen-1 G3 fragments, particularly for the shorter G3-II and G3-III fragments (Fig. 1a, b). The YWTD β -propeller domain alone is sufficient to form a stable complex with LE3-5 (Fig. 1b, lane 6). A second nidogen present in vertebrates, nidogen-2 (ref. 3), gave similar results (Fig. 1b, lanes 7 and 8). This redundancy in laminin binding is consistent with the observation that mice deficient in nidogen-1 or nidogen-2 are mildly or not overtly affected¹⁵⁻¹⁷. The nidogen-1 complex with LE3-5 is stable to purification and storage in physiological buffer. However, the complex dissociates at pH 6 and below (Fig. 1c), in marked contrast to the low-density lipoprotein receptor (LDLR), in which association of the β -propeller domain with LA modules 4 and 5 requires acidic pH (ref. 10).

A crystal structure of the nidogen-1 G3-III complex with laminin LE3-5 was solved by using molecular replacement to a resolution of 2.3 Å (Fig. 2a, Supplementary Table 1). As predicted¹¹, the nidogen G3 domain assumes a symmetrical six-bladed β -propeller fold. Each β -sheet or 'blade' of the propeller contains four antiparallel β -strands, with β -strand 1 lining the solvent-filled central channel around the pseudo-6-fold symmetry axis, and β -strand 4 outermost and forming the cylindrical side of the propeller (Fig. 2a). The loops are named after the strands they connect; the 4-1 loops and 2-3 loops alternate in anticlockwise order on the 'top' face, whereas the 1-2 and 3-4 loops form the flared 'bottom' face where the central

channel is wider. There is an offset of the six YWTD repeats with respect to the blades, so that the three amino-terminal β -strands unite with the carboxy-terminal β -strand to form blade 6, closing the circular topology of the propeller (Fig. 3). Nidogen contains a long-range disulphide bond that connects blade 1 to blade 6 at the bottom of the β -propeller and is unique among YWTD domains (Figs 2a and 3). The long-range disulphide bond is located at a force-bearing point where neighbouring domains connect, and would help to prevent the β -propeller from unravelling when the basement membrane is mechanically stressed.

The structure of LE3-5 bound to nidogen is similar to that of uncomplexed LE3-5 (refs 18, 19), except for a bend in LE3,

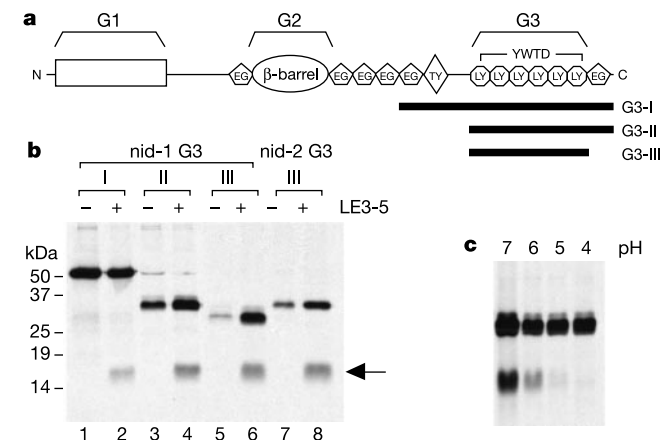


Figure 1 Nidogen and laminin fragment complexes. **a**, Nidogen-1 fragments. EG, EGF module; TY, thyroglobulin repeat; LY, LDL receptor YWTD repeat. **b**, Co-expression of nidogen and laminin γ 1 fragments. Material was immunoprecipitated from [³⁵S]methionine and cysteine-labelled culture supernatants with antibody against the V5 tag (Invitrogen) on the nidogen fragment and Protein G-agarose, and subjected to reducing SDS-polyacrylamide-gel electrophoresis. Co-precipitated LE3-5 is marked with an arrow. **c**, Nidogen-laminin association is pH dependent. nidG3-III/LE3-5 immunoprecipitates were incubated at the indicated pH at room temperature for 30 min and washed; material remaining bound was eluted with SDS sample buffer.

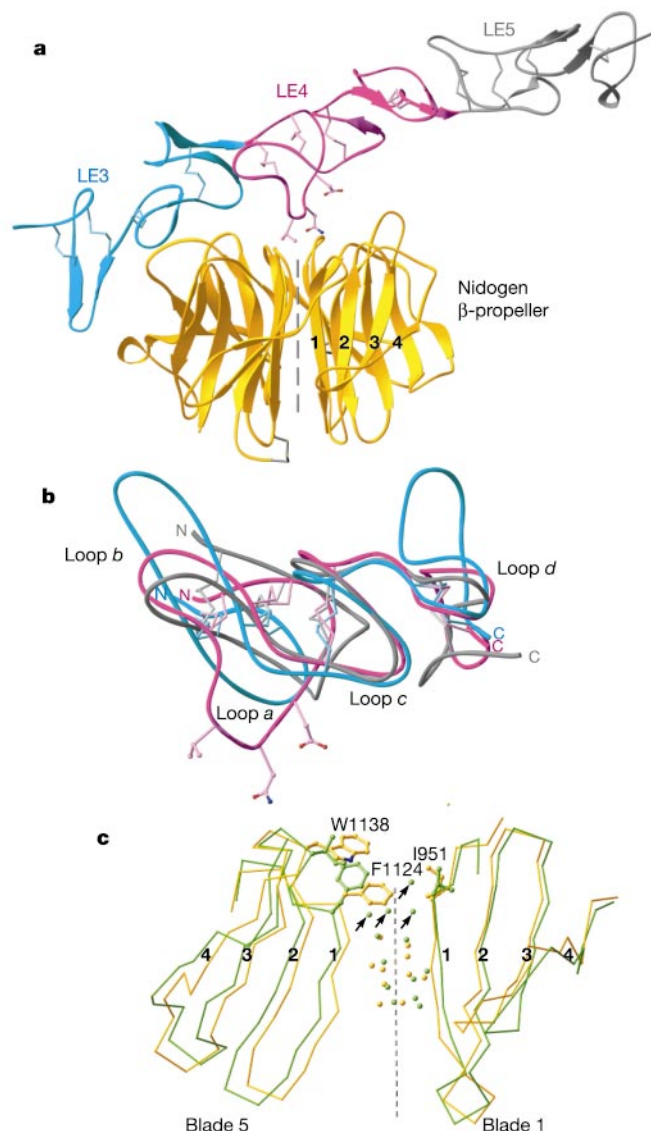


Figure 2 The nidogen β -propeller complex with laminin modules LE3-5 and comparison with the LDLR β -propeller. **a**, Ribbon diagram of the complex. β -strands are numbered on one β -propeller β -sheet. The two disulphide connections are black. **b**, Superposition of LE modules 3 (cyan), 4 (magenta), and 5 (grey). Disulphide connections are thin bonds in the same colour; Asp 800, Asn 802 and Val 804 side chains of LE4 are shown in **a** and **b**. **c**, Superposition of nidogen and LDLR β -propellers showing β -sheets 1 and 5. The C α trace backbones, side chains and water molecules in the central channel are shown in gold (nidogen) and green (LDLR). Side chains in the hydrophobic shutter in nidogen and equivalent residues in LDLR are shown. The four water molecules present in LDLR and absent in nidogen are marked with arrows. Dashed lines in **a** and **c** represent the pseudo-6-fold axis.

confirming that intermodule interactions give tandem LE modules a rigid rod-like structure (Supplementary Fig. 1A). LE domains are a variant of the EGF fold, with an extra disulphide-bonded *d* loop²⁰. The longer disulphide-bonded loops have been designated the *a*, *b*, *c* and *d* loops¹⁸ (Fig. 2b).

LE module 4 docks to an amphitheatre-shaped concave surface centred on the pseudo-6-fold axis atop the β -propeller, whereas LE3 protrudes over the rim of the propeller and exclusively contacts blade 6 (Figs 2a and 3). The unique *a* loop of LE4 (Fig. 2b) forms a prominent, ridge-like protrusion into the amphitheatre of the β -propeller (Fig. 2a). LE3 and LE4 bury similar amounts of total solvent-accessible surface area of 1,000 and 870 Å², respectively. However, the shape complementarity²¹ of 0.82 at the LE4 interface is much higher, and is also higher than for previously tabulated subunit or receptor–ligand interfaces²¹. In agreement, studies with recombinant LE modules¹² show that the LE4 module makes the predominant contribution to binding affinity, and LE3 is insufficient for binding but augments affinity. LE4 binds with an affinity of 10⁹ M⁻¹ or more (refs 12–14) yet buries only 870 Å², making it the smallest high-affinity interface yet known⁵. The unusual imbalance in burial of 520 Å² on LE4 and 350 Å² on nidogen reflects the marked convexity and concavity of their binding surfaces, respectively⁵.

Rigidity minimizes loss of entropy upon binding and promotes high affinity. On nidogen, rigidity is provided by the concave, amphitheatre-shaped interface (Fig. 4a). The residues in the β -bulge of β -strand 1 (b position; Fig. 3) form the innermost, lower tier of seats in the amphitheatre. Many of these residues point

upwards and contact laminin (Figs 3 and 4a), whereas Ile 951 and Phe 1124 point inwards towards the pseudo-symmetry axis and form a hydrophobic shutter that seals off the floor of the amphitheatre (Fig. 4a), shielding it from the water-filled central aperture of the β -propeller that lies just below the binding pocket (Fig. 2c, Supplementary Fig. 1b). The residues immediately preceding β -strand 1 and following β -strand 2, in the 0 position of the β -ladder (Fig. 3), form an outer, higher tier with twice as many seats. These residues point upwards and also form many of the laminin contacts (Fig. 4a). The amphitheatre is strongly rigidified by within-sheet backbone hydrogen bonds, and by between-sheet hydrogen bonds from the side chains of the Thr and Asp of the YWTD motif to backbone. The constraints imposed by packing the bulky side chains that line the amphitheatre onto a concave surface also conspire to create a rigid binding surface.

The rigidity of the binding surface on laminin is shown by the remarkable similarity of LE4 loop *a* and its side chains in the uncomplexed and complexed LE3–5 structures¹⁸ (Supplementary Fig. 1a). The LE4 loop *a* backbone is well supported by hydrogen bonds (Supplementary Fig. 3), the importance of which is emphasized by the conservation in metazoan evolution of Asp 774 and Asn 806 (Supplementary Fig. 2), which form hydrogen bonds to the loop *a* backbone. Moreover, mutation of loop *c* residues Ile 818 and Tyr 819, which do not interact with nidogen but stabilize loop *a* conformation, reduces the affinity for nidogen by 170-fold and 25-fold, respectively¹⁴.

Interactions within the binding site seem highly efficient. Asn 802 and Val 804 in loop *a* of LE4 protrude most deeply into the propeller amphitheatre, and are centred on the pseudo-6-fold axis (Figs 2a and 4a). The 3,000-fold loss of affinity upon mutation Val 804 → Ser (ref. 14) demonstrates the importance of its hydrophobic interaction with multiple β -propeller amphitheatre residues. The side chain of Asn 802 of LE4 forms two complementary hydrogen bonds to the side chain of Asn 1082 of nidogen (Fig. 4a). The importance of both hydrogen bonds and the stereochemistry of this interaction is demonstrated by the decrease in affinity to 1/5,000 or less after mutation of Asn 802 to Asp, Gln or Ser (ref. 14). Furthermore, the backbone carbonyl of Asn 802 forms bidentate hydrogen bonds to Arg 1037 of nidogen, which also hydrogen bonds with the Asn 802 side chain (Fig. 4a). These interactions are further stabilized by the presence of Arg 1037 at the centre of an extensive network of hydrogen bonds (Supplementary Fig. 3). The side chains of residues Asp 800 of LE4 and Arg 1055 of nidogen form a salt bridge through two charged hydrogen bonds (Fig. 4a). Mutations Asp 800 → Asn and Asp 800 → Glu reduce affinity by 8,000-fold and 70-fold, respectively, confirming the importance of the salt bridge¹⁴. The importance of residues 800, 802 and 804 agrees with the ability of the *a* loop heptapeptide spanning residues 798–804 to bind to nidogen with an affinity of 10⁶ M⁻¹ (ref. 13).

In comparison with other extracellular protein recognition interfaces, that between nidogen and laminin module LE4 is unusually conserved. On nidogen, interacting Glu, Asn, Leu, Tyr, two Arg and two Trp residues are invariant in mammalian nidogen-1 and nidogen-2 and in the *C. elegans* and *Drosophila* nidogens, and Val and Ile residues are conserved as aliphatics (Fig. 3). On LE4, Asp 800, Asn 802 and Val 804 are conserved in mammalian γ 1 and γ 3 laminin subunits, and in the single *Drosophila* and *C. elegans* γ subunits, with the exception of a conservative Val → Ile substitution in *C. elegans* (Supplementary Fig. 2). Notably, *Drosophila* laminin binds to mammalian nidogen²². The extreme conservation of the nidogen–LE4 interface suggests that it is highly optimized structurally.

In contrast to nidogen, all other proteins that contain YWTD β -propeller domains are cell surface receptors¹¹. Many function as endocytic receptors, such as the LDLR^{11,23}, and many also function in developmental pathways, including photoreceptor cell development in *Drosophila*, head regeneration in *Hydra*, and neuronal positioning and bone development in vertebrates^{24,25}. The recent

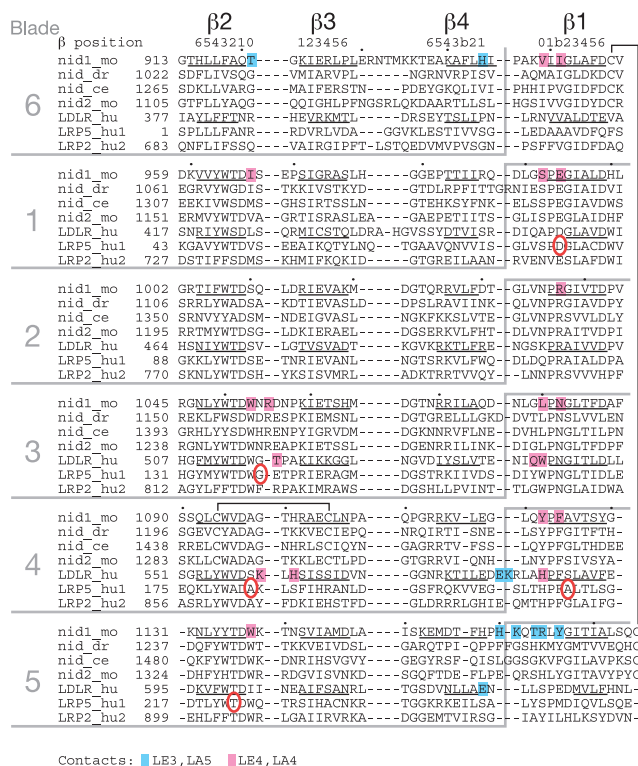


Figure 3 Sequence alignment. Structural alignment of nidogen-1 and LDL receptor²³ β -propeller domains and alignment by sequence to other YWTD domains. mo, mouse; dr, *Drosophila*; ce, *C. elegans*; hu, human; the 1 and 2 suffixes refer to the first and second β -propellers. Each row corresponds to one YWTD repeat; the offset blade boundaries are indicated by grey lines. Blades are structurally aligned, and equivalent β -ladder positions in each strand are numbered at the top of the alignment. β -strands are underlined. Residues in LRP5 known to cause unregulated Wnt signalling in patients with increased bone mass when mutated^{6,7,9} are denoted by a red circle; two positions have two independent substitutions each.

structure of the extracellular fragment of the LDLR at endosomal pH shows that the lipoprotein-binding LDLR type A (LA) modules 4 and 5 loop back and bind to the YWTD β -propeller domain, suggesting a function in receptor recycling¹⁰. Histidines have central positions in the intramolecular LDLR interface but not in the nidogen–laminin interface, which is consistent with the distinct pH preferences of these interactions. The LA4 and LA5 modules of the LDLR also bind lipoprotein ligands, in agreement with a role of the β -propeller domain in ligand release and receptor recycling at endosomal pH. LA4 binds to the top face of the β -propeller using

evolutionarily conserved residues¹⁰, similar to the interaction of LE4 with nidogen. However, in contrast to LE4, LA4 has no loop that inserts into the hub of the β -propeller (Fig. 4b), and LA4 binds off-centre (Fig. 4c). Whereas LE4 contacts all six blades of the nidogen β -propeller, LA4 contacts only blades 3, 4 and 5 (Figs 3 and 4c). Like LA5, LE3 binds over the rim of the β -propeller (Fig. 4b, c). The highest projection from the rim is formed by the 2–3 loop of blade 3, which interacts with both LE4 and LA4 (Fig. 3), whereas the shortest 2–3 loop is in blade 6, near the LE3 and LA5 interaction sites. This asymmetric length of loops on opposite sides of the β -propeller rim is

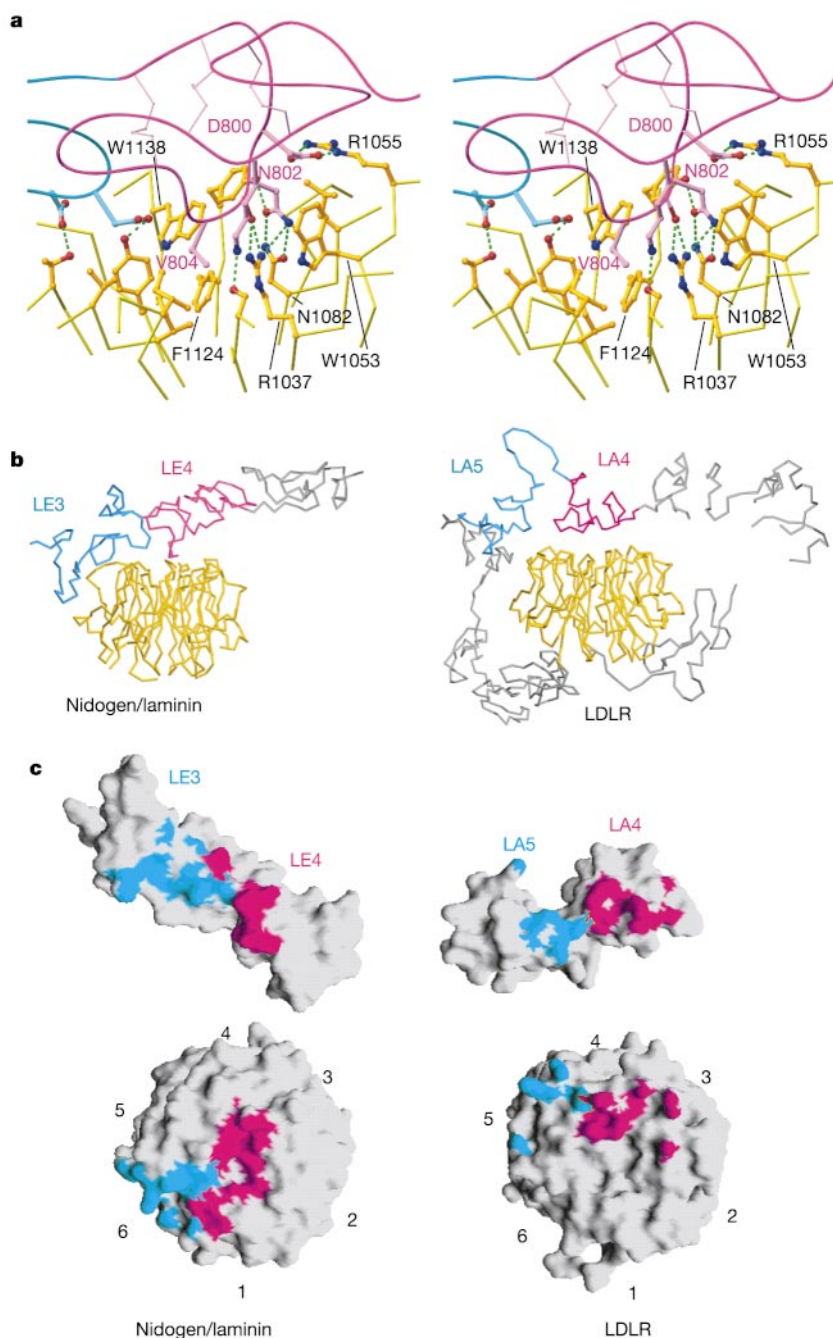


Figure 4 The nidogen–laminin interaction and comparison with the interaction in the LDLR. **a**, Stereo view of the nidogen interaction with LE4 and the adjacent portion of LE3. LE3 and LE4 are as in Fig. 2a. Portions of the nidogen backbone as C α -trace and side chains forming the amphitheatre are shown in gold. Green dashed lines represent hydrogen bonds. **b**, C α trace representation of the nidogen–laminin and intrinsic LDLR interactions. The propeller axes are vertical in the page. For clarity, the views are towards

β -propeller blade 1 in nidogen and blade 6 in LDLR. **c**, GRASP surfaces. Atoms within 4 Å of an interaction partner are shown in magenta (LE4 and LA4) and cyan (LE3 and LA5). The β -propeller domains are viewed down their 6-fold axes. In the open book representation, the interacting LE3–4 or LA4–5 module pairs are rotated through 180° in the horizontal plane of the page. β -propeller blades are numbered.

characteristic of YWTD β -propellers (Fig. 3), and correlates with the slanted orientation with which both the LE and LA modules bind (Fig. 4b).

Remarkably, the LDLR YWTD β -propeller domain differs from that of nidogen in lacking closure of the bottom of the binding pocket in the hub of the β -propeller. The 1.5-Å crystal structure of the LDLR β -propeller module²³ reveals a continuous column of water that extends all the way through the central cavity (Fig. 2c). In nidogen, Phe 1124 in β -strand 1 of blade 5 acts as a shutter to close the central aperture, whereas the equivalent Phe residue in the LDLR points upward and does not block the aperture (Fig. 2c, Supplementary Fig. 1b). The key difference is that in the upper, outer tier of residues in blade 5, Trp 1138 in the 2–3 loop of nidogen substitutes for an Ile in the LDLR (Fig. 3). The bulky Trp 1138 in nidogen pushes Phe 1124 into the central aperture, ensuring the hydrophobicity of the LE4-binding pocket (Fig. 2c). Furthermore, in the nidogen β -propeller, blades 1 and 5 are closer to the pseudo-6-fold axis by 0.8 and 1 Å, respectively than in either LDLR structure^{10,23} (Fig. 2c). Together with the Phe 1124 shutter, this closer packing excludes four water molecules from the nidogen β -propeller that are present at the narrow, upper portion of the central channel in the LDLR, and ensure a hydrophobic binding pocket with a low dielectric constant. The hydrophobic shutter in nidogen seems important for its high affinity for laminin. Because in the LDLR the LA4 and LA5 modules are covalently tethered to the β -propeller domain, high affinity is not required for their interaction.

The concept that YWTD β -propeller domains can bind extrinsic ligands with high affinity, and can also bind intrinsic ligands and promote receptor recycling after endocytosis, has important implications for the increasingly large family of receptors that contain YWTD domains and function in regulating development as well as in endocytosis^{24,25}. Many ligands may bind to YWTD domains rather than LA modules, and binding to YWTD domains might affect signalling directly, enable binding to co-receptors, antagonize binding to intrinsic LA module ligands, or affect receptor endocytosis and recycling. Of particular relevance are LRP5, LRP6 and *Drosophila* Arrow, each of which has four YWTD domains, acts as a Wnt co-receptor and also binds with nanomolar affinity to cysteine-rich domains in Dickkopf proteins. Mutations in LRP5 that eliminate its expression lead to reduced bone mass in humans⁸. By contrast, a mutant LRP5 that causes high bone mass^{6,9} remains activatable by Wnt but not its antagonist Dickkopf-1 (ref. 9). Remarkably, of seven independent hypermorphic amino acid substitutions in LRP5, all map to the first YWTD β -propeller domain, in positions that are structurally equivalent to the binding site in nidogen for LE4 (Fig. 3). The mutations in YWTD β -propeller 1 of LRP5 all lie in the 4–1 and 2–3 loops of blades 2, 3, 4 and 5. In the number of blades included, the participation of residues that are deep in the central amphitheatre such as Asp 80 (Fig. 3), the predicted presence of a Phe shutter, and high affinity for ligand, the binding modality of LRP5 more resembles that of nidogen than the LDLR β -propeller domain. These results suggest that the high-affinity ligand-binding site described here may be paradigmatic for signalling receptors that bear YWTD domains, as well as of importance in the formation of basement membranes. □

Methods

Expression constructs for nidogen (entactin) fragments contained the authentic signal sequence and various segments from the G3 domain, followed by C-terminal V5 epitope and hexahistidine tags. Nidogen-1 fragments were G3-I, mature amino acid residues 774–1217; G3-II, 913–1217; and G3-III, 913–1179. The nidogen-2 G3-III fragment contained residues 1105–1373. Mouse entactin complementary DNA (a gift from A. E. Chung) or mouse entactin-2 cDNA (a gift from M. Shimane)² were used as PCR templates. The mouse laminin γ 1 chain encompassing LE3–5 (residues 736–899) was amplified from mouse spleen cDNA library and cloned downstream of a signal sequence without any tag. Complex between nidG3-III and LE3–5 was transiently expressed in 293-EBNA human kidney cells in serum-free medium using episomal expression vectors pCEP4 (Invitrogen) and pCEP-Pu (a gift from R. Timpl)²⁶ and purified from culture supernatant by Ni-affinity chromatography

after dialysis against 25 mM Tris-HCl pH 8.0, 150 mM NaCl, and eluted with 250 mM imidazole in the same buffer. The complex (~7 mg) was treated with 10 μ g ml⁻¹ chymotrypsin at room temperature for 2 h to remove the flexible tag segment, further purified by Mono Q chromatography (50 mM Tris-HCl pH 8.5, eluted with a NaCl gradient), dialysed against 10 mM Tris-HCl pH 7.5 containing 10 mM NaCl, and concentrated.

Crystals were grown in hanging drops at room temperature by mixing equal volumes of protein (13 mg ml⁻¹) and reservoir solution (1 M sodium acetate, 0.05 M CdSO₄, 0.1 M HEPES pH 7.5). The crystal was transferred to reservoir solution supplemented with 20% (v/v) glycerol and flash-frozen in liquid nitrogen.

Data were collected on beamline 19-ID at the Advanced Photon Source at Argonne National Laboratory (Argonne, Illinois), and processed with program HKL2000 (ref. 27). The structure was solved with molecular replacement using program AmoRe²⁸ with a two-body search in the 15–3.5-Å range with models of the LDLR YWTD domain²³ and LE3–5 (ref. 18). The rotation function peaks for LE3–5 stood out significantly, and the correct one for nidogen was the third highest. A significant translation function was first found for LE3–5, and with LE3–5 fixed, a highly significant translation function was found for nidogen. There is one complex per asymmetric unit. Rigid body refinement yielded a clean map in which main-chain and side-chain density was easily seen for model rebuilding with O. The model was refined to 2.3 Å resolution using simulated annealing in CNS version 1.1 (ref. 29). Six Cd²⁺ ions are bound to the complex. Two interact with the β -propeller only. Four are at boundaries between nidogen and LE3–5 and to some extent are involved in crystal packing, but do not have a specific role in complex formation, as confirmed by lack of effect of EDTA on complex stability (data not shown).

For β -propeller blade superpositions, blades from nidogen and LDLR²³ were cut out and all 12 were simultaneously structurally aligned using 3DALIGN of Modeller 6.0 (ref. 30). For β -propeller superpositions, all equivalent residues determined by blade superposition were used.

Received 10 March; accepted 26 June 2003; doi:10.1038/nature01873.

- Hynes, R. O. & Zhao, Q. The evolution of cell adhesion. *J. Cell Biol.* **150**, F89–F96 (2000).
- Mayer, U., Kohfeldt, E. & Timpl, R. Structural and genetic analysis of laminin–nidogen interaction. *Ann. N.Y. Acad. Sci.* **857**, 130–142 (1998).
- Timpl, R. in *Guidebook to the Extracellular Matrix, Anchor, and Adhesion Proteins* (eds Kreis, T. & Vale, R. D.) 455–457 (Oxford Univ. Press, 1999).
- Willems, M. *et al.* Specific ablation of the nidogen-binding site in the laminin γ 1 chain interferes with kidney and lung development. *Development* **129**, 2711–2722 (2002).
- Lo Conte, L., Chothia, C. & Janin, J. The atomic structure of protein–protein recognition sites. *J. Mol. Biol.* **285**, 2177–2198 (1999).
- Little, R. D. *et al.* A mutation in the LDL receptor-related protein 5 gene results in the autosomal dominant high-bone-mass trait. *Am. J. Hum. Genet.* **70**, 11–19 (2002).
- Van Wesenbeeck, L. *et al.* Six novel missense mutations in the LDL receptor-related protein 5 (LRP5) gene in different conditions with an increased bone density. *Am. J. Hum. Genet.* **72**, 763–771 (2003).
- Gong, Y. *et al.* LDL receptor-related protein 5 (LRP5) affects bone accrual and eye development. *Cell* **107**, 513–523 (2001).
- Boyd, L. M. *et al.* High bone density due to a mutation in LDL-receptor-related protein 5. *N. Engl. J. Med.* **346**, 1513–1521 (2002).
- Rudenko, G. *et al.* Structure of the LDL receptor extracellular domain at endosomal pH. *Science* **298**, 2353–2358 (2002).
- Springer, T. A. An extracellular β -propeller module predicted in lipoprotein and scavenger receptors, tyrosine kinases, epidermal growth factor precursor, and extracellular matrix components. *J. Mol. Biol.* **283**, 837–862 (1998).
- Mayer, U. *et al.* A single EGF-like motif of laminin is responsible for high affinity nidogen binding. *EMBO J.* **12**, 1879–1885 (1993).
- Poschl, E., Fox, J. W., Block, D., Mayer, U. & Timpl, R. Two non-contiguous regions contribute to nidogen binding to a single EGF-like motif of the laminin gamma 1 chain. *EMBO J.* **13**, 3741–3747 (1994).
- Poschl, E. *et al.* Site-directed mutagenesis and structural interpretation of the nidogen binding site of the laminin γ 1 chain. *EMBO J.* **15**, 5154–5159 (1996).
- Murshed, M. *et al.* The absence of nidogen 1 does not affect murine basement membrane formation. *Mol. Cell Biol.* **20**, 7007–7012 (2000).
- Dong, L. *et al.* Neurologic defects and selective disruption of basement membranes in mice lacking entactin-1/nidogen-1. *Lab. Invest.* **82**, 1617–1630 (2002).
- Schymeinsky, J. *et al.* Gene structure and functional analysis of the mouse nidogen-2 gene: Nidogen-2 is not essential for basement membrane formation in mice. *Mol. Cell Biol.* **22**, 6820–6830 (2002).
- Stetefeld, J., Mayer, U., Timpl, R. & Huber, R. Crystal structure of three consecutive laminin-type epidermal growth factor-like (LE) modules of laminin γ 1 chain harboring the nidogen binding site. *J. Mol. Biol.* **257**, 644–657 (1996).
- Baumgartner, R. *et al.* Structure of the nidogen binding LE module of the laminin γ 1 chain in solution. *J. Mol. Biol.* **257**, 658–668 (1996).
- Bork, P., Downing, A. K., Kieffer, B. & Campbell, I. D. Structure and distribution of modules in extracellular proteins. *Q. Rev. Biophys.* **29**, 119–167 (1996).
- Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950 (1993).
- Mayer, U., Mann, K., Fessler, L. I., Fessler, J. H. & Timpl, R. *Drosophila* laminin binds to mammalian nidogen and to heparan sulfate proteoglycan. *Eur. J. Biochem.* **245**, 745–750 (1997).
- Jeon, H. *et al.* Implications for familial hypercholesterolemia from structure of the LDL receptor YWTD-EGF domain pair. *Nature Struct. Biol.* **8**, 499–504 (2001).
- Herz, J. & Bock, H. H. Lipoprotein receptors in the nervous system. *Annu. Rev. Biochem.* **71**, 405–434 (2002).
- Strickland, D. K., Gonias, S. L. & Argraves, W. S. Diverse roles for the LDL receptor family. *Trends Endocrinol. Metab.* **13**, 66–74 (2002).

26. Kohfeldt, E., Maurer, P., Vannahme, C. & Timpl, R. Properties of the extracellular calcium binding module of the proteoglycan testican. *FEBS Lett.* **414**, 557–561 (1997).
27. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
28. Navaza, J. Amore: An automated package for molecular replacement. *Acta Crystallogr. A* **50**, 157–163 (1994).
29. Brunger, A. T. *et al.* Crystallography & NMR System: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
30. Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank R. Zhang and A. Joachimiak for beamline help, and S. Blacklow and R. Timpl for critically reading the manuscript. This work was supported by an NIH grant.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.A.S. (springero@med.harvard.edu) or J.W. (jwang@red.dci.harvard.edu). The PDB code for the nidogen/laminin complex is 1NPE.

corrigendum

¹⁴⁶Sm–¹⁴²Nd evidence from Isua metamorphosed sediments for early differentiation of the Earth's mantle

Guillaume Caro, Bernard Bourdon, Jean-Louis Birck & Stephen Moorbath

Nature **423**, 428–432 (2003).

In this Letter, the subscripts *l* and *k* were accidentally transposed, so coefficient *A* in equation (4) should have read:

$$A = \frac{\ln(m_i/m_j) \ln(m_i/m_l)}{\ln(m_k/m_j) \ln(m_k/m_l)} \quad \square$$

errata

Links between signal transduction, transcription and adhesion in epithelial bud development

Colin Jamora, Ramanuj DasGupta, Pawel Kocieniewski & Elaine Fuchs

Nature **422**, 317–322 (2003).

In Figs 1e and 2j, k of this paper, the colour keys are incorrect. In Fig. 1e, the blue bars indicate TOPFLASH and the purple bars indicate FOPFLASH. The description of the results in the text is correct. In Fig. 2j and k, the green fluorescence indicates P-cadherin (Pcad) and E-cadherin (Ecad), respectively, and the red fluorescence indicates the marker for the basement membrane, laminin (lam). □

Class 3 semaphorins control vascular morphogenesis by inhibiting integrin function

Guido Serini, Donatella Valdembri, Sara Zanivan, Giulia Morterra, Constanze Burkhardt, Francesca Caccavari, Luca Zammataro, Luca Primo, Luca Tamagnone, Malcolm Logan, Marc Tessier-Lavigne, Masahiko Taniguchi, Andreas W. Püschel & Federico Bussolino

Nature **424**, 391–397 (2003).

In Figs 3c, g and 4c, g of this Article, the colour keys are incorrect. The light grey and white boxes were interchanged. The description of the results in the text is correct. □

Antibodies in abundance

The latest offerings for immunology.

ELISA kits for human sIL-2R and IL-2

Bender MedSystems www.instant-elisa.com

Assay in four steps

An instant ELISA kit for detecting the soluble form of human interleukin-2 receptor (sIL-2R) and interleukin-2 (IL-2) is available from Bender MedSystems. IL-2R has been shown to provide a useful and simple means of assessing immune function *in vivo*. Elevated levels of soluble IL-2R have been associated with a broad spectrum of human diseases, including many cancers, autoimmune and inflammatory diseases, infectious diseases, organ rejection, impaired renal function and thermal injuries. IL-2 has been shown to play a central role in immune responsiveness and provides insights into several pathological processes, including cancer. The instant ELISA kits can be used with a wide variety of biological sample types. The assay procedure is a simple four-step process: add water and sample; incubate and wash; add TMB; and stop the reaction. Each kit includes a microtitre plate containing freeze-dried, pre-mixed quantities of the standard protein, coating antibody, streptavidin HRP and sample diluent.

Immun-Star HRP

Bio-Rad www.bio-rad.com

A little light relief

Bio-Rad has expanded its Immun-Star line of chemiluminescent kits for western blot detection to include a new kit for horseradish peroxidase (HRP) conjugated secondary antibodies. This highly sensitive, fast-acting alternative to isotopic detection can detect less than one picogram of sample, and the light signal lasts for 6–8 hours after activation, so there is no need to rush into the darkroom immediately. Film exposures are as little as 30 seconds to 2 minutes on commonly used

blotting supports. Most blots can be reused by stripping and reprobing. Working solutions are stable at room temperature for 24 hours, and the shelf-life of the kits is one year at 4°C. The kits can also be used with imaging systems such as Bio-Rad's ChemiDoc and VersaDoc, and are available as substrate only, substrate with conjugate or substrate with conjugate and blotting reagents.

Product guide

Axxora www.axxora.com

Reagents for B-cell signalling

A full-colour product guide describing innovative reagents for B-cell signalling studies is now available from Axxora. Antibodies, ligands, receptors and ELISA kits for CD40L, BAFF and APRIL are among the products listed. Product descriptions include complete technical data as well as selected literature references and review articles. Breakdowns in B-cell signalling processes have been associated with severe malignancies, such as non-Hodgkin's lymphoma, and autoimmune diseases, including systemic lupus erythematosus and rheumatoid arthritis. The guide can be ordered by visiting www.alexis-e.biz, and is free of charge.

Claudin antibodies

Zymed www.zymed.com

Negotiate tight junctions

Claudins are a novel family of integral membrane proteins that were first identified as essential elements of the tight junction following observations that the disruption of both occluding alleles did not prevent the assembly of functional tight junction strands. Each member of the claudin family bears four transmembrane domains and is unrelated by sequence to occludin. At least 18 homologous proteins have been identified in humans. Claudins are important structural and functional components of tight junctions in paracellular transport. They are located in both epithelial and endothelial cells in all tissues that bear tight junctions and occur in a complex with occluding and junctional adhesion molecules. Claudin defects have been implicated in a wide range of human disease, with claudins 1, 2, 3, and 5 associated with colon cancer, claudin 3 with prostate cancer and claudin 4 with pancreatic cancer. Zymed offers claudins 1, 2, 3, 4 and 5 for use in formalin-fixed, paraffin-embedded tissues in immunohistochemistry. The company also supplies a wide range of claudin antibodies for use in western blotting, immunofluorescence and immunoprecipitation.



Star turn: a new detection kit from Bio-Rad.

FACS Sample Prep Assistant

BD Biosciences www.bdbiosciences.com

Automate routine tasks in the lab

The FACS Sample Prep Assistant can be used along with the BD FACS Lyse/Wash Assistant to automate sample preparation for flow cytometry in clinical and research applications. Both instruments free laboratory personnel from time-consuming sample preparation tasks, reduce potential exposure to biohazards, provide flexibility with customizable protocols and process up to 40 samples at a time. The Sample Prep Assistant automatically mixes primary blood collection tubes, aliquots blood from primary to staining tubes, dispenses antibody reagents into staining tubes and mixes, adds lysing solution to staining tubes and mixes, and times the staining and lyse incubations.

Montage antibody purification kits

Millipore www.millipore.com

Purification by centrifugation

Millipore's Montage antibody purification kits with PROSEP-A and PROSEP-G media can be used to achieve fast and convenient antibody purification by centrifugation. The kits include ready-to-use spin columns prepacked with immobilized protein A or G media for antibody purification from serum, ascites or cell culture supernatants. The centrifugal protocols are easy to follow and allow walk-away convenience. Media plugs are prepacked. The protocols purify 15–25 mg of antibody in less than one hour.

RNAiFect transfection reagent

Qiagen www.qiagen.com

Silence, please

Transfection of siRNA is a critical step in gene silencing experiments. RNAiFect transfection reagent, based on a lipid formulation, provides efficient transfection of siRNA into a wide range of eukaryotic cells. Its low endotoxin levels and absence of RNase activity ensure optimal performance in demanding RNAi experiments. The reagent can be used for transfection of a wide range of cell types, including

AGS, HEK-293, HeLaS3, Huh-7, HUVEC, NIH-3T3 and PA317 cells. RNAiFect reagent is provided in a ready-to-use form for easy transfection—simply add the reagent to the diluted siRNA, mix, incubate, and pipette the complexes onto the cells. Transfection can be performed in the presence of serum, eliminating the need to remove the complexes from the cells. Formation of transfection complexes occurs efficiently in complete growth medium. In some cell lines, the silencing effect is significantly enhanced by replacing growth medium with Buffer EC-R during complex formation. RNAiFect reagent is also available in bulk for high-throughput experiments.

Negative isolation kit

Dynal Biotech www.dynalbiotech.com

Untouched monocytes

The monocyte negative isolation kit (now available in a new 25 ml volume) with Dynal's MPC-50 magnet allow users to work with 15–50 ml samples and isolate from up to 5×10^8 mononuclear cells (MNC). The improved antibody mix contains CD235a antibody, which removes erythrocytes present in the sample, thereby improving purity. The untouched monocytes are suitable for deriving dendritic cells. The gentle, tube-based method allows isolation from prepared MNC in 45 minutes. Negative isolation uses the principle of magnetic separation, removing unwanted components in a mixed starting sample to leave viable and untouched target cells in the supernatant. Each kit contains Depletion Dynabeads and an optimized antibody mix. Kits are also available in a 5 ml format and in formulations suitable for other cell types, including T cells, CD4 and CD8 T cell subsets, B cells and NK cells.

Streptamers

IBA www.iba-go.com

Isolated T cells

IBA's Streptamer technology is designed for isolation and characterization of functional, antigen-specific T cells. It provides reversible, antigen-specific staining, leaving the isolated T cells phenotypically and functionally

indistinguishable from untreated cells. Other features include high staining intensities, mild and rapid dissociation of staining reagents using biotin and improved viability of T cells after isolation. Several human (HLA) and mouse (H2) MHC 1-*Strep* alleles with viral or cancer antigens are available.

DD-4X

Genevac

www.genevac.co.uk

Versatile workhorse evaporator

The DD-4X centrifugal evaporator from Genevac is a solvent evaporation system that combines ease of use with sampling versatility and enhanced productivity. For drug discovery researchers faced with handling samples in a variety of formats, the DD-4X accommodates a wide range of tubes, vials, microtitre plates and Fast-Stack multi-deck microplate holders. Other features include a cost-effective diaphragm pump, a redesigned evaporation chamber and powerful infrared lamps for efficient removal of solvents. Effective chamber cooling between runs protects thermally sensitive samples, which can be run immediately after higher boiling point solvents. High condensing power and a large-capacity waste solvent cold trap maximize solvent recovery while minimizing both run and condenser defrost times.

Celera Discovery system

Applied Biosystems

www.appliedbiosystems.com

Web-based discovery platform expanded

Applied Biosystems has released an enhanced version of the Celera Discovery system (CDS) online platform, which now includes an individual subscription model for academic researchers. Subscribers have online access to an integrated set of data and analysis tools. The new release has an interactive map viewer with customizable graphic displays for comparative genomics studies. Researchers can also conduct BLAST searches in batch mode and customize data views, from entire chromosomes down to single nucleotide level. Other features include curated human and mouse gene databases, updated sequence analysis tools, protein libraries for functional predictions and links to the Human Gene Mutation Database (HGMD), Online Mendelian Inheritance in Man (OMIM), and the Swiss-Prot and LocusLink databases.

AS laser microdissection instrument

Leica Microsystems

www.leica-microsystems.com

A cut above the rest

Leica's AS laser microdissection instrument (LMD) can be used for isolation of single cells using high-resolution microscopy and immunofluorescence. The laser system is attached to an upright microscope with bright field and epifluorescence, enabling single cells to be accurately identified and cut out. Larger sub-structures in a tissue, such as follicles of lymph nodes, can also be isolated. Host-pathogen interactions can be investigated at the molecular level, and changes in gene expression in individual cells during disease pathogenesis can be observed. Current applications include studies of pathogens that affect farm animals, including the viruses that cause foot-and-mouth disease and classical swine fever, and bacteria such as *Salmonella* and *Mycobacterium bovis*, which causes bovine tuberculosis.

ProtoCOL

Synbiosis

www.synbiosis.com

Rapid quality assurance for makers of vaccines

ProtoCOL is a reaction zone reader and colony counter that provides an indication of vaccine potency by automatically measuring reaction zones in single radial immunodiffusion (SRD) assays. The system consists of a CCD camera integrated with a PC and software. It will automatically read an entire SRD plate, measure reaction zones and transcribe the results in Excel. The software also flags areas of dispute such as fuzzy edges, so that users can measure them manually with the on-screen ruler. The flexible Excel format enables users to export



Determining vaccine potency with ProtoCOL.

the results into most statistical programs without re-keying. Synbiosis claims that the ProtoCOL system has the potential to increase throughput by at least a hundredfold.

CosMCPrep

Agencourt Bioscience

www.agencourt.com

Shotgun sequencing

The CosMCPrep plasmid purification system uses a single protocol to purify a variety of high- and low-copy number template types, including BACs, fosmids and cosmids. With increasing reliance on whole-genome shotgun sequencing, fosmid end sequencing has become a necessity for every genome project. Traditionally, purification of low-copy plasmids can be difficult due to the low numbers per cell. CosMCPrep eliminates this problem, enabling efficient, fully automated purification of low-copy plasmids.

Immobilizer microarray slide

Flowgen

www.flowgen.co.uk

Hold still

Flowgen's Immobilizer is an open, ready-to-spot polymer microarray slide with a transparent spotting area that is coated with anthraquinone photo probes to enable one-step coupling (spotting) of amino-linked DNA/RNA, -LNA or -PCR. In between the spots is low nonspecific binding of DNA and fluorochrome labelled proteins. The slide requires no activation steps, blocking, baking or ultraviolet crosslinking, and is supplied with spotting and hybridization solutions. The 75 × 25 × 1 mm slide dimension makes it compatible with most microarray spotter and scanners. Finger indentations are made to facilitate manual handling, and an elevated bottom minimizes scratches. Slides are stable at room temperature for at least 12 months and can be bar-coded for data tracking. Applications include clone mapping, linkage analysis, highly multiplexed mutation detection, gene expression monitoring and high-throughput screening.

These notes are compiled in the Nature office from information provided by the manufacturers.

Contacts

Publisher: Ben Crowe
Editor: Paul Smaglik
Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Global Head of Naturejobs:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)
Andy Douglas (4975)
Frank Phelan (4944)

Netherlands/ Italy/ Iberia/ Belgium:

Evelina Rubio Hakansson (4973)
Scandinavia: Silje Opstrup (4994)

France/ Switzerland:

Amelie Pequignot (4974)
Natureevents:
Paul Constant (4954)

Production Manager:

Billie Franklin
To send materials use London
address above.
Tel +44 (0) 20 7843 4814
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

International

Advertising Coordinator:
Hind Berrada (4935)

Naturejobs web development:
Tom Hancock

Naturejobs online production:
Ben Lund

European Satellite Office

Germany/ Austria:

Patrick Phelan, Odo Wulffen
Tel + 49 89 54 90 57 11/2
Fax + 49 89 54 90 57 20
e-mail: p.phelan@nature.com
o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@nature.com

US Sales Manager: Peter Bless

US Advertising Coordinator:
Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Haraikatamachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746

Asia-Pacific Sales Director:
Rinoko Asami
e-mail: r.asami@nature.com

naturejobs

Fears for foreign physicists

In the years since the terrorist attacks of 11 September 2001, there has been anecdotal evidence that young foreign scientists have become more hesitant about studying and working in the United States. Tougher visa rules announced earlier this year seemed to discourage them further (see *Nature* **422**, 96–97; 2003). A survey by the American Institute of Physics (AIP) now shows that those rules, and perhaps the general climate of suspicion in the United States towards foreign scientists, has led to a reduction in the number of young physicists entering the country.

The proportion of foreign physics students in the United States had risen steadily since 1971, when 21% of all first-year graduate students came from abroad. It reached a peak of 55% for the 2000–01 academic year, but has since dropped by 10%.

The AIP survey shows that visa difficulties have played a large part in the decline. In 2002, about 20% of foreign applicants accepted to physics graduate programmes in the United States were initially denied entry into the country. The group experiencing the biggest problems hasn't been from the Middle East, but rather from China, which provides the largest number of foreign physics graduate students.

Visa problems haven't yet discouraged foreign students from applying, but they probably will unless the rules are relaxed. And if the decline in foreign students continues, it could have profound effects on physics employment in the United States. Many foreign physicists stay on past their training, so there may be fewer talented people to fill physics posts after graduate school.

This may well lead to more opportunities for those students who do make it into the US physics programme. But the drop in competition and international collaboration means that those who find these jobs could be celebrating bitter-sweet victories.

Paul Smaglik

Naturejobs editor



Contents

MOVERS

Graduate institute gets fresh blood; astrophysicist takes control; new head welcome at centre for life; and more **p980**

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

Transitions

X Peter Lee was this month appointed chief executive of AgriGenesis, a plant-science firm based in Auckland, New Zealand.

X Richard Gibbs, director of the Human Genome Sequencing Center at Baylor College of Medicine, **Jorge Galán**, professor and chairman of microbial pathogenesis at the Yale University School of Medicine, and **Ian Lipkin**, epidemiology professor at the College of Physicians and Surgeons, Columbia University, were this month appointed to the scientific advisory board of 454 Life Sciences, a genomic-sequencing-technology company in Branford, Connecticut.

X Michael Valentino was this month appointed president and chief executive of Adams Laboratories, a pharmaceutical company based in Fort Worth, Texas, that specializes in drugs for respiratory care. Valentino, who takes over the reins from **Mark Gainor**, was president and chief operating officer, global pharmaceuticals, of Alpharma, a generic-drug company based in New Jersey.

X James Rothman was last month appointed chief scientific adviser for Amersham. Rothman, currently chairman of the Cellular Biochemistry and Biophysics Program at the Memorial Sloan-Kettering Cancer Center in New York, was recently recruited to Columbia University to found the Center for Chemical Biology.

X Zahed Subhan has joined Copenhagen-based biotech company Nuevolution as chief executive. Subhan was previously vice-president of business development at Locust Pharmaceuticals, a drug-development company that is based in Pennsylvania.

LIFE SCIENCES

Last month, **Sheldon Schuster** became the new president of the Keck Graduate Institute (KGI) in Claremont, California, close to Los Angeles. He succeeds the KGI's founding president, **Henry Riggs**. Riggs set up the graduate educational and research institution, which is focused on applied life sciences, in 1997 to break down barriers between academia and industry. There are 17 faculty members.

Schuster brings considerable experience in higher education and research to the job. Before joining the KGI, he was assistant vice-president for research and graduate education, and director of the biotechnology programme, at the University of Florida in Gainesville. He was also a professor of biochemistry and molecular biology with an interest in studying enzyme targets for developing anti-cancer drugs. "I have had experience in all of the pieces of this and was just excited about the opportunity to see it brought together," Schuster says.

PHYSICAL SCIENCES

Astrophysicist **Michael Turner** of the University of Chicago is to head the mathematical and physical sciences directorate at the US National Science Foundation (NSF). The directorate, which commands 20% (or US\$1 billion for the current fiscal year) of the NSF's total budget, supports research in mathematics, physics, chemistry, materials and astronomy, as well as multidisciplinary and educational programmes.

Turner takes up the post on 1 October, initially serving for a two-year term. But he also intends to continue with his research, dividing his time between the NSF and the University of Chicago, where he is chair of the department of astronomy and astrophysics. Turner's research career spans three decades. He is a member of the US National Academy of Sciences and is also a senior scientist at the Fermi National Accelerator Laboratory in Illinois.

"I think the strongest credential that I bring to this job is my close connection with the scientific community," he says. Great attention has been paid to the biological sciences over the past decade, he adds; this is an opportunity to redress the balance and "really move the physical sciences forward".

As assistant director, Turner will set the vision for the directorate — one of seven at the NSF — and must make the case before Congress. He would, for instance, like to see greater support for research probing the origin and evolution of the Universe.

Another area would be high-energy



Sheldon Schuster



Michael Turner



Michael Dexter



Bernice Welles

physics, which, in the United States, "is not healthy right now", says Turner.

GENETICS

Michael Dexter, former director of the Wellcome Trust, will head the International Centre for Life in Newcastle, UK, starting in November. Dexter headed the Wellcome Trust during the international Human Genome Project, which culminated in a freely available draft sequence in February 2001. At the Centre for Life he will help steer the Wellcome-supported institution on its mission to engage the public in science — especially in the areas of genetics and genomics. Dexter will succeed outgoing chairman **Matt Ridley**.

BIOTECHNOLOGY

Bernice Welles, who has a degree in music, a graduate degree in urban and environmental policy, and a medical degree, has followed a similarly varied professional track. This month she joined the MPM BioVentures group as a venture partner in its San Francisco office. Previously Welles worked as an independent biotech consultant, spent almost eight years at biotech company Genentech, most recently as vice-president of product development, and started out on her postgraduate career as an assistant professor in the Department of Medicine at the University of California, San Francisco.

Welles says that her diverse educational and professional paths are signs of the times, with people making more changes in their career and making education a perpetual goal, rather than a finite destination. For Welles, that flexibility meant following seemingly unrelated interests that eventually complemented one another.

As well as following her loves, she also used some pragmatism. She liked her music-theory classes so much that she considered doing a PhD in the subject, but graduate students in it convinced her otherwise. "Most of them were unemployed," Welles says. "The lucky ones were writing jingles."

So she hedged her bets by also pursuing pre-med coursework. That turned out to be a prudent decision, but the mathematical aspects of music theory helped her, both in her pursuit of her public-health-related master's and now in her analysis of investment portfolios. Her permanent education continues, with coursework towards an executive MBA under way. "I am kind of an education junkie," Welles says. "But I think this will be my last degree."

CONTACTING US AT MOVERS

Please send any information on appointments or job moves to:
naturejobseditor
@naturedc.com